numero 115
marzo 1978
anno XI
volume XX

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

# Il problema dell'anidride carbonica

*Il consumo di combustibili fossili e la distruzione delle foreste hanno determinato un aumento di anidride carbonica nell'atmosfera, il quale potrebbe portare a profonde modificazioni climatiche*

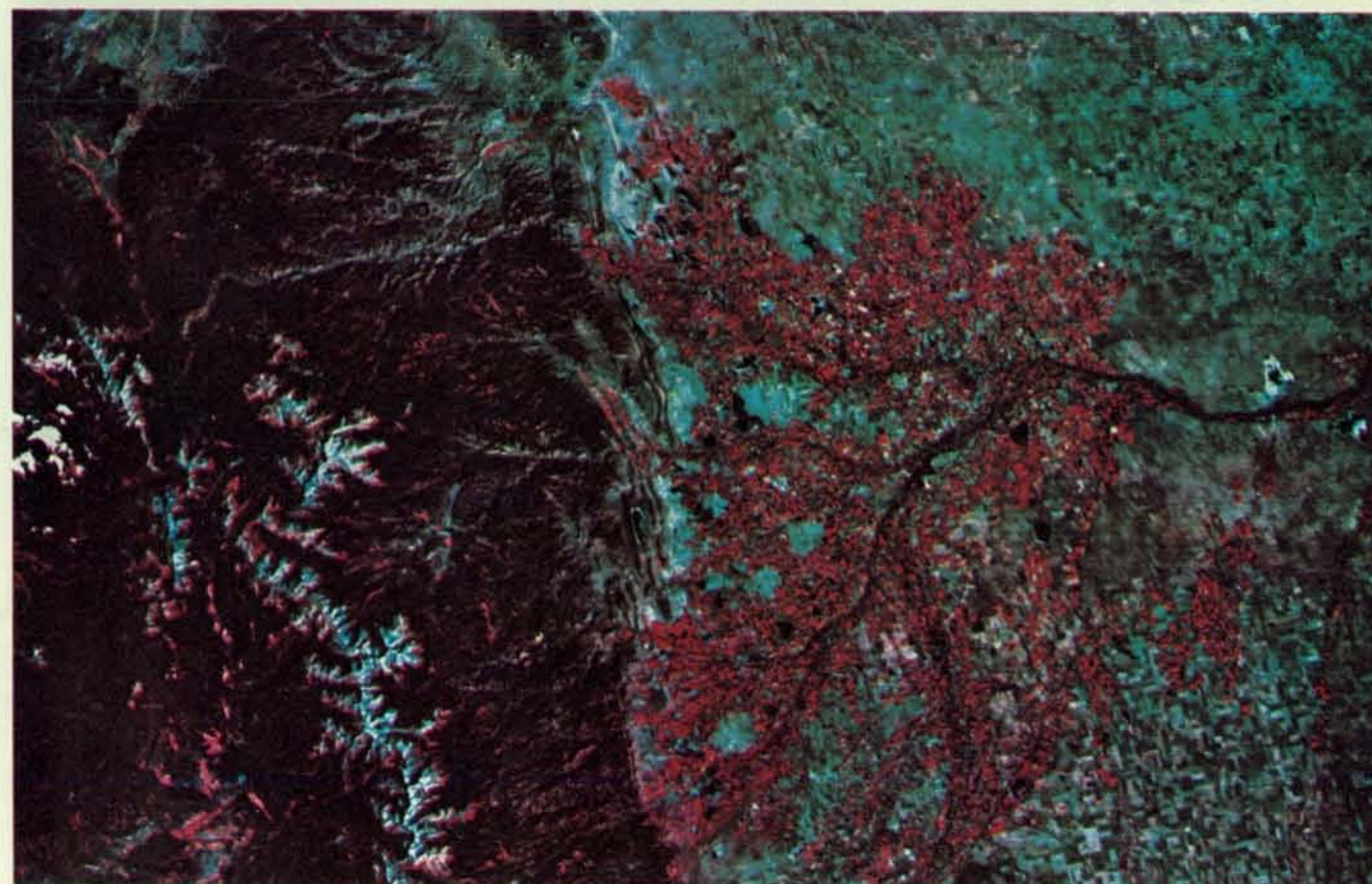
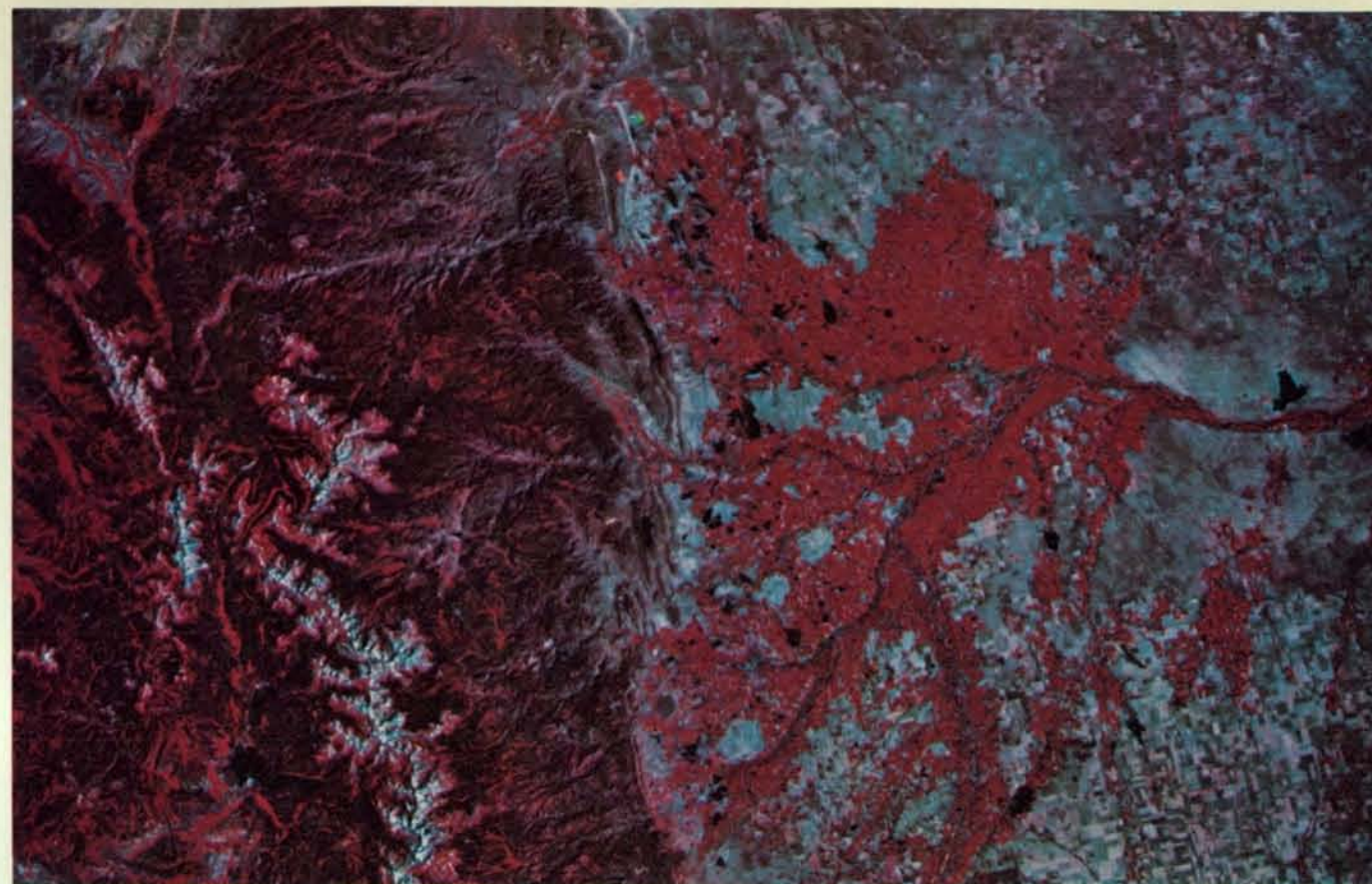
di George M. Woodwell

Dal 1850 a oggi, le attività umane hanno prodotto un aumento nel contenuto atmosferico di anidride carbonica da 290 parti per milione, o anche meno, a poco più di 330 parti per milione. Forse un quarto dell'aumento totale si è avuto nel corso dell'ultimo decennio. Permanendo le tendenze attuali, entro l'anno 2020 il contenuto atmosferico di anidride carbonica potrebbe arrivare a sfiorare il raddoppio rispetto ai valori presenti. Fino a epoca recente, tale incremento veniva concordemente attribuito alla combustione di combustibili fossili, ma attualmente è emersa la possibilità che a esso contribuisca in egual misura un'altra circostanza: la distruzione di foreste su scala planetaria.

Anche se l'anidride carbonica è un gas presente solo in tracce nell'atmosfera terrestre - la sua concentrazione è infatti circa dello 0,3 per cento in volume - esso esercita probabilmente un'influenza critica sul controllo del clima terrestre, poiché assorbe energia radiante alle lunghezze d'onda dell'infrarosso. Il calore che in questo modo viene intrappolato è potenzialmente in grado di alterare profondamente il clima terrestre. Inoltre, affatto indipendentemente dai possibili effetti che può esercitare sul clima, l'anidride carbonica ha un'importanza critica anche come fonte del carbonio che viene fissato per fotosintesi dalle piante verdi e costituisce la base dell'esistenza del mondo animale e vegetale.

L'umanità si trova quindi di fronte a un dilemma storico. Le attività umane, che stanno producendo un continuo aumento del contenuto atmosferico di anidride carbonica, lasciano presupporre un generale riscaldamento del clima nel corso dei decenni a venire. Anche se non si conoscono con certezza né l'entità dei cambiamenti climatici che interverranno né il preciso meccanismo attraverso cui avranno luogo, gli effetti connessi a una crescita costante della quantità di anidride carbonica nell'atmosfera produrranno quasi sicuramente squilibri. Un aumento nelle temperature medie della Terra si tradurrà probabilmente in un ampliamento delle zone aride e in sostanziali ripercussioni sulla produzione agricola.

L'altro corno del dilemma sta nel fatto che il genere di azioni correttive che potrebbero essere prese in considerazione avrebbero sicuramente conseguenze egualmente destabilizzanti. La più ovvia delle azioni correttive consisterebbe in una più drastica riduzione del consumo di combustibili fossili. Uguale importanza rivestirebbero misure volte a rallentare il ritmo con cui le foreste della Terra vengono ridotte o trasformate in radura in seguito a sfruttamento del legname, ad ampliamento dei terreni agricoli e da pascolo, a inquinamento e ad altre conseguenze dello sviluppo industriale. Un più energico sforzo per modificare la ripartizione tra l'uso agricolo e forestale delle estensioni di terreno in concomitanza con uno sforzo per limitare il consumo di combustibili fossili finirebbe per sconvolgere a tal punto i modelli consolidati di sviluppo sociale ed economico, da produrre nelle condizioni umane alterazioni altrettanto drastiche di quelle probabilmente conseguenti a un riscaldamento del clima.

Benché il problema dell'anidride carbonica esista da oltre un secolo, è solo dal 1958 che si è entrati in possesso di dati inequivocabili circa le variazioni del contenuto atmosferico di questo composto. In quell'anno, Charles D. Keeling della Scripps Institution of Oceanography creò una stazione di rilevamento continuo dell'anidride carbonica sul vulcano Mauna Loa, nell'isola Hawaii. Il Mauna Loa fu scelto perché offriva la possibilità di studiare il contenuto di anidride carbonica nella troposfera, cioè nella parte inferiore dell'atmosfera, alle medie latitudini. Le registrazioni ora disponibili provenienti da Mauna Loa e da altre stazioni mostrano chiaramente due ordini di regolarità. In primo luogo, durante il periodo successivo al 1958 si è manifestata una regolare tendenza all'aumento del contenuto di anidride carbonica. Il valore dell'aumento a Mauna Loa è stato di circa 8 parti per milione per anno, anche se sono ovviamente intervenute variazioni nel tasso di tale incremento. In secondo luogo, si osserva una oscillazione sistematica del contenuto atmosferico di anidride carbonica in correlazione con l'avvicendamento delle stagioni. Il tenore di anidride carbonica tocca una punta massima in inverno inoltrato, che generalmente cade in aprile nell'emisfero settentrionale, e precipita a

**L'annuale variazione della fissazione di anidride carbonica, caratteristica delle zone temperate, viene mostrata nelle immagini (riprese in estate e in autunno dal satellite Landsat) di una regione lungo il margine orientale delle Montagne Rocciose vicino a Boulder, Colorado. In questo tipo di fotografie, il verde della vegetazione appare rosso, Nell'immagine in alto, effettuata in agosto, il rosso intenso dell'area a destra delle montagne, sulla sinistra, denuncia un massimo di attività fotosintetica, durante la quale le piante, per alimentarsi, assorbono anidride carbonica dall'atmosfera. Nell'immagine in basso, ripresa in ottobre, il rosso si è sostanzialmente attenuato, e ciò denota un rallentamento nell'attività fotosintetica. L'attività fotosintetica qui considerata si riferisce tanto a foreste quanto a piante coltivate.**

un minimo verso la fine dell'estate settentrionale, a settembre inoltrato o a ottobre. I dati di Mauna Loa costituiscono la più lunga e accurata registrazione continuata del contenuto di anidride carbonica mai prima d'ora effettuata in alcuna parte del mondo.

Registrazioni della concentrazione di anidride carbonica sono state effettuate, in corrispondenza a vari periodi, al Polo Sud, in Australia, a Point Barrow in Alaska, a Long Island (New York) e in altre località. Ricercatori hanno eseguito estese campionature dell'atmosfera negli Stati Uniti, in Svezia, Australia e altrove a bordo di aerei. Tutti i dati mostrano un'oscillazione inverno-estate, con un minimo nella tarda estate e un massimo nel tardo inverno. Le oscillazioni seguono le stagioni di ciascun emisfero. I dati mostrano inoltre un aumento più o meno continuo del tenore atmosferico di anidride carbonica e l'entità di questo aumento varia, in funzione dell'epoca e della località, da circa 0,5 a 1,5 parti per milione per anno.

La variazione stagionale nella concentrazione atmosferica di anidride carbonica riflette uno dei principali fattori che influenzano l'atmosfera: il metabolismo del biota, ossia della totalità della materia vivente. Tale variazione stagionale è in correlazione con l'incremento della fotosintesi, che si verifica durante l'estate alle medie latitudini di entrambi gli emisferi. Recentemente è stata accettata come valida l'ipotesi che la causa primaria della variazione stagionale è, con tutta probabilità, la variazione della fotosintesi nelle foreste delle medie latitudini. Viene attribuita importanza alle foreste perché, oltre a coprire estese superfici, svolgono su scala planetaria più attività fotosintetica di ogni altro tipo di vegetazione e sono in grado di immagazzinare carbonio in quantità sufficiente per influenzare il contenuto atmosferico di anidride carbonica.

Le variazioni di ampiezza nella differenza tra le concentrazioni invernali e quelle tardo-estive di anidride carbonica sono coerenti con questa ipotesi. La differenza varia da circa cinque parti per milione a Mauna Loa a più di quindici parti per milione nel centro di Long Island. La differenza si attenua verso i tropici, dove la fluttuazione stagionale del metabolismo è meno pronunciata oppure assente; la differenza si indebolisce anche con l'elevarsi dell'altitudine a tutte le latitudini. L'ampiezza si riduce sostanzialmente nell'emisfero meridionale, apparentemente a causa della minore estensione di foreste conseguente alla più limitata massa di terre continentali. La conclusione cui chiaramente si giunge è che le foreste della Terra esercitano una spiccata influenza sull'andamento a breve termine della concentrazione di anidride carbonica nell'atmosfera.

La causa dell'aumento a lungo termine del contenuto atmosferico di anidride carbonica - aumento stimabile nell'ordine del 10-15 per cento dal 1850 - è stato generalmente attribuito alla immissione, in continuo aumento, nell'aria di anidride carbonica attraverso il consumo di combustibili fossili. Recenti analisi hanno gettato dubbi su questa interpretazione. La probabilità che esista una rilevante introduzione addizionale di anidride carbonica nell'atmosfera a opera del biota è stata messa in luce da me insieme ad alcuni miei colleghi dell'Ecosystems Center del Marine Biological Laboratory a Woods Hole, Mass., in collaborazione con R.H. Whittaker e Gene E. Likens della Cornell University, W.A. Reiners del Darmouth College e C.C. Delwiche dell'Università della California a Davis. Analoghe interpretazioni sono state pubblicate da altri ricercatori, specialmente da Bert Bolin dell'Università di Stoccolma e da J.R. Adams e dai suoi colleghi della Rice University. L'emissione a opera del biota è avvenuta principalmente in seguito alla distruzione di foreste e all'ossidazione dell'humus. La supposizione che l'aumento del contenuto atmosferico di anidride carbonica sia stata una conseguenza del consumo dei combustibili fossili, senza tener conto di possibili modificazioni nel biota, ha condotto a ciò che ora appare come un grave errore di calcolo nella costruzione del bilancio terrestre del carbonio.

L'inconveniente ha tratto origine dalle supposizioni che il biota abbia agito come un serbatoio di accumulo dell'anidride carbonica atmosferica, mentre in realtà esso si è probabilmente comportato come una fonte di liberazione nell'atmosfera di anidride carbonica. L'esistenza di questo probabile errore compromette al presente la possibilità di risolvere i principali quesiti inerenti al bilancio terrestre del carbonio. Se il biota non è stato un serbatoio di accumulo dell'anidride carbonica atmosferica e se l'assorbimento della stessa a opera degli oceani della Terra non è superiore a quanto si è supposto, allora la quantità di anidride carbonica presente nell'atmosfera dovrebbe essere in aumento anche più velocemente di quanto mostrano le osservazioni. Ovviamente le stime sono sbagliate. Ma dove sta l'errore?

Il problema può essere colto più chiaramente se si paragonano le grandezze dei «depositi» tra i quali avviene un reciproco scambio di carbonio in modo più o meno continuo. L'atmosfera contiene attualmente circa $700 \times 10^{15}$ grammi di carbonio sotto forma di anidride carbonica la quale viene continuamente scambiata con il biota e con le acque di superficie degli oceani. La quantità di carbonio depositato nel biota terrestre complessivo ammonta a circa $800 \times 10^{15}$ grammi, cioè è leggermente superiore alla quantità presente nell'atmosfera. Una quantità ancora maggiore di carbonio, variamente valutata tra i $1000 \times 10^{15}$ e $3000 \times 10^{15}$ grammi, è trattenuta nei materiali organici del suolo, principalmente humus e torba. Lo sfruttamento delle foreste, l'estendersi dell'agricoltura su terreni che contengono grandi quantità di materiale organico e la distruzione di ambienti umidi sono tutti fattori che accelerano la degradazione dell'humus, il quale viene trasformato in anidride carbonica, acqua e calore. L'anidride carbonica liberata entra nel «deposito» atmosferico.

**L'area di foresta pluviale nel bacino amazzoniano del Brasile nordoccidentale che appare in questa immagine, ripresa dal Landsat, è quasi uniformemente rossa, a dimostrazione dell'intensa attività fotosintetica, lungo l'arco annuale, che caratterizza le foreste pluviali tropicali. Le comunità delle foreste naturali fissano più carbonio per unità di superficie che non la maggior parte delle piantagioni agricole. L'estensione di queste aree forestali si sta tuttavia riducendo.**

**L'andamento del contenuto atmosferico di anidride carbonica è stato misurato dal 1958 all'osservatorio di Manua Loa, nell'isola Hawaii, da Charles D. Keeling della Scripps Institution of Oceanography. I puntini indicano la concentrazione media mensile. Le oscillazioni stagionali sono provocate dalla sottrazione di anidride carbonica per fotosintesi durante la stagione della crescita nell'emisfero settentrionale e dalla successiva liberazione durante l'autunno e l'inverno. Le misurazioni effettuate a Manua Loa e altrove mostrano che il contenuto medio di anidride carbonica nell'atmosfera è salito di più del 5 per cento dal 1958. La velocità di incremento è variata di anno in anno per cause non note. La velocità attuale è di una parte per milione per anno, equivalente a $2,3 \times 10^{15}$ grammi di carbonio.**

Rispetto a questi tre depositi in continua interazione e di dimensioni approssimativamente uguali, gli oceani raccolgono complessivamente una quantità di carbonio molto maggiore. La maggior parte del carbonio si trova sotto forma di anidride carbonica disciolta, che è parte del sistema carbonati-bicarbonati. Includendo i grandi bacini oceanici, la quantità totale di carbonio accumulata in questa forma è di $40\,000 \times 10^{15}$ grammi. Rispetto a una scala temporale misurata in migliaia di anni, il contenuto atmosferico di anidride carbonica potrebbe a ben ragione essere determinato dall'equilibrio stabilitosi con il carbonio inorganico delle acque profonde degli oceani. La velocità di scambio tra l'atmosfera e gli oceani nel loro complesso è tuttavia bassa. Gli scambi più rapidi intervengono tra l'atmosfera e lo strato superficiale di mescolamento, dello spessore approssimativo di 100 metri, che sovrasta le più fredde acque abissali. Lo strato superficiale contiene circa $600 \times 10^{15}$ grammi di carbonio inorganico. Il deposito oceanico di carbonio secondo per vastità si trova sotto forma di materia organica disciolta (l'«humus» degli oceani) che sembra essere presente dovunque in ragione di una parte per milione, totalizzando una quantità che probabilmente arriva a $3000 \times 10^{15}$ grammi considerando tutti gli oceani. Le acque profonde degli oceani superano ampiamente per volume le acque dello strato di mescolamento e racchiudono il deposito di carbonio di gran lunga più abbondante che sia in rapporto di scambio con l'atmosfera: da $35\,000 \times 10^{15}$ a $38\,000 \times 10^{15}$ grammi. (Non viene qui considerato il carbonio dei sedimenti carbonacei che ne contengono quantità assai più rilevanti.) La capacità delle regioni abissali di assorbire carbonio è virtualmente illimitata. Il fattore limitante è dovuto all'estrema lentezza con cui sembra avvenire il passaggio del carbonio dell'atmosfera verso le acque profonde attraverso lo strato superficiale degli oceani.

Quando si cerca di costruire un diagramma di flusso che indichi il trasporto netto di carbonio da un serbatoio all'altro, si scopre che le stime disponibili presentano ampi scostamenti qualitativi. I dati più accurati di cui si è in possesso si riferiscono all'anidride carbonica liberata sulla Terra attraverso il consumo di combustibili fossili, che attualmente equivale a $5 \times 10^{15}$ grammi di carbonio ogni anno e all'aumento del contenuto atmosferico di anidride carbonica equivalente a circa $2,3 \times 10^{15}$ grammi di carbonio per anno. Ci sono quindi $2,7 \times 10^{15}$ grammi di carbonio proveniente da combustibili fossili che devono venire allontanati da una qualche combinazione di processi terrestri e oceanici. Limitiamoci a supporre, per il momento, che il biota terrestre rappresenti un serbatoio stabile di carbonio, cioè che non agisca nel senso di aumentare o di diminuire la quantità di anidride carbonica presente nell'atmosfera. In questo caso (di cui sosterrò l'improbabilità) gli oceani sono chiamati ad accogliere $2,7 \times 10^{15}$ grammi di carbonio ogni anno. La domanda che ci si deve porre è se questo tasso di assorbimento sia dimostrabile.

Secondo i calcoli più attendibili degli oceanografi chimici è difficile spiegare come possano gli oceani assorbire una tale quantità annualmente. Le loro analisi si basano su una conoscenza dettagliata della quantità di carbonio interessata nel sistema carbonati-bicarbonati dello strato superficiale e sulla costruzione di accurati modelli del meccanismo di mescolamento degli oceani. Gli isotopi radioattivi carbonio 14 e trizio (idrogeno 3), entrambi prodotti in grandi quantità durante gli esperimenti atomici degli anni cinquanta e dell'inizio degli anni sessanta, sono stati impiegati come traccianti per studiare la velocità con cui l'acqua dello strato di mescolamento si scambia con l'acqua degli strati profondi. Gli

studi sembrano indicare che la velocità di mescolamento è in verità molto bassa. Il trasferimento del carbonio dall'atmosfera allo strato superficiale e da questo alle acque abissali ha poca probabilità di superare, secondo alcuni colleghi, i $2,5 \times 10^{15}$ grammi per anno. In conclusione, gli oceani appaiono appena sufficienti a fungere da vaso di raccolta della differenza tra i $5 \times 10^{15}$ grammi per anno di carbonio attualmente liberati nell'atmosfera bruciando combustibili fossili e i $2,3 \times 10^{15}$ grammi che l'atmosfera trattiene.

Questo bilancio di flussi deve essere completamente riesaminato se il serbatoio biotico di carbonio, anziché espandersi o essere in equilibrio netto con il carbonio dell'atmosfera, è in realtà una fonte netta di anidride carbonica per la atmosfera. Whittaker e Likens hanno recentemente ordinato in tabelle le informazioni attuali circa la dimensione dei diversi segmenti del biota. Dal loro lavoro risulta che i più estesi serbatoi di carbonio in seno al biota sono le foreste. Risulta inoltre che la massima produzione fotosintetica si svolge sulle terre emerse e non negli oceani come prima era stato supposto sulla base di precedenti argomentazioni intorno alla velocità della produzione primaria netta negli oceani. (Per produzione primaria netta si intende la quantità netta residuale di carbonio fissato per fotosintesi, cioè di materia organica, una volta che sono state soddisfatte le necessità respiratorie della pianta. Si tratta della materia organica utilizzabile per la crescita della pianta che viene poi immagazzinata o resa disponibile per gli organismi consumatori.)

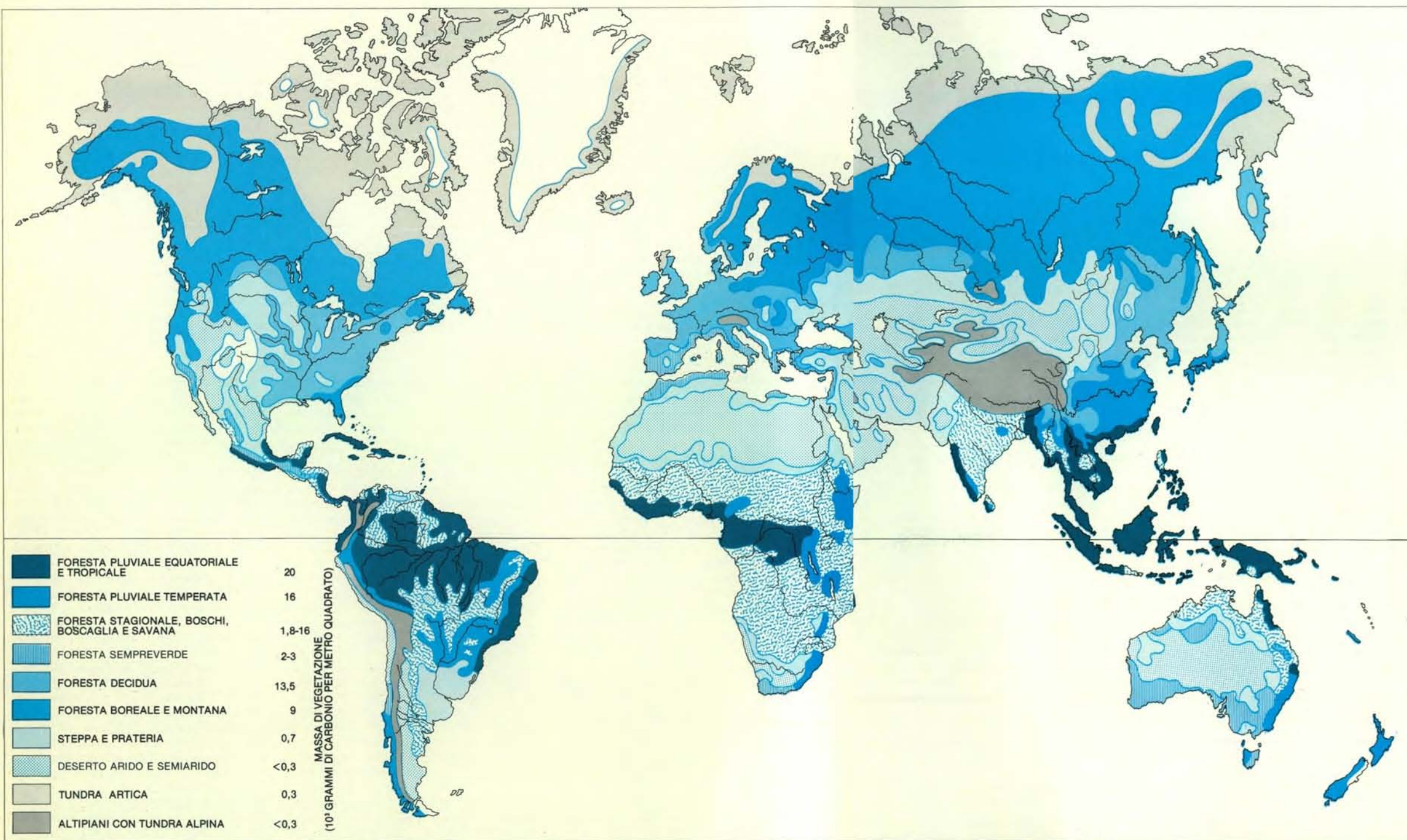
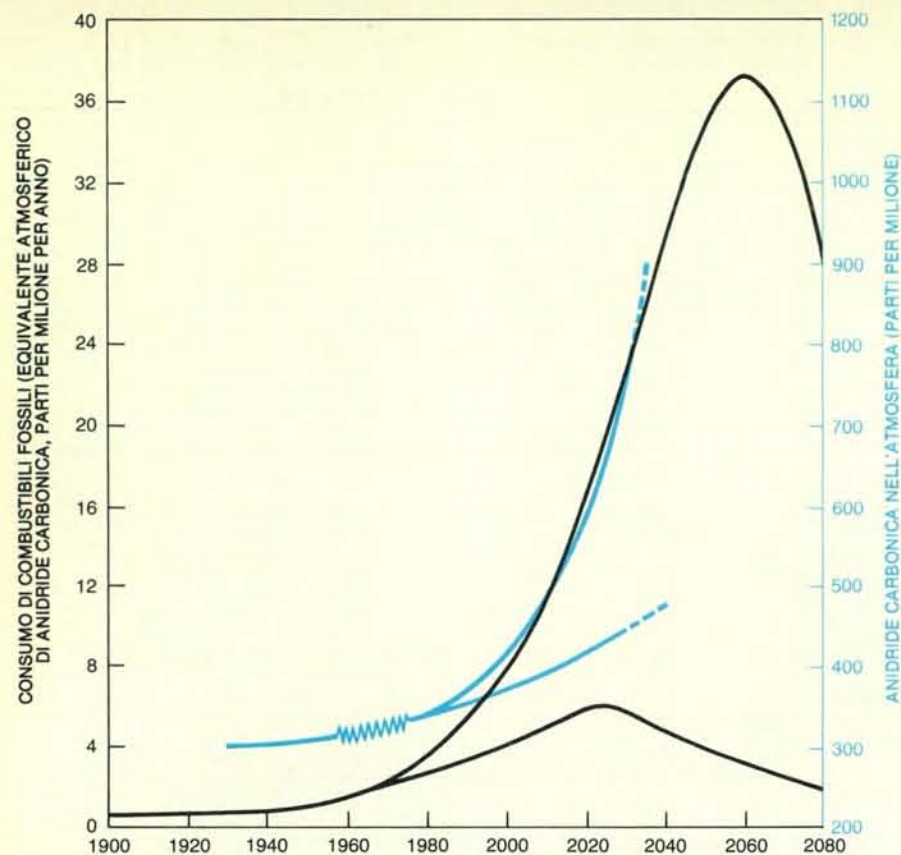
Forse il risultato più significativo dello studio di Whittaker e Likens è la dimostrazione che le foreste pluviali tropicali, con i loro grandi alberi, rappresentano il più grande serbatoio singolo di carbonio in seno al biota, oltre a detenere il primato assoluto della produzione totale primaria netta. Questa considerazione sottolinea l'importanza delle foreste tropicali nel bilancio terrestre del carbonio. Se queste e altre foreste verranno tagliate rapidamente e se il carbonio in esse contenuto verrà liberato, esiste la concreta possibilità che venga significativamente influenzata la quantità di anidride carbonica nell'atmosfera. Per contro, se verrà permesso il rimboschimento dei terreni disboscati, le foreste assorbiranno parte dell'anidride carbonica dall'atmosfera rallentando l'aumento della sua concentrazione.

Curiosamente, solo in epoca recente sono sorte controversie circa il quesito se il serbatoio biotico si stia ingrandendo o rimpicciolendo. Per esempio, durante una conferenza nel 1970 intitolata «Study of Critical Environmental Problems» organizzata dal Massachussets Institute of Technology e tenutasi a Williamstown, Mass., i relatori sostennero che il serbatoio biotico doveva essere in fase di espansione e incoraggiavano gli oceanografi a reputare adeguati i loro modelli di circolazione oceanica e di assorbimento dell'anidride carbonica.

La prima seria sfida al modello degli oceanografi fu lanciata attraverso due monografie presentate alla Conferenza Dahlem sulla biogeochimica tenutasi a Berlino Ovest nel novembre 1976. In una monografia, R.A. Houghton e io valutammo che il biota probabilmente libera ogni anno tanta anidride carbonica nell'atmosfera quanto quella prodotta nella combustione dei combustibili fossili. Nell'altra monografia, Bolin giunse a una valutazione leggermente inferiore per quanto riguarda il biota, prendendo spunto da dati sullo sfruttamento forestale raccolti dalla Food and Agriculture Organization (FAO) delle Nazioni Unite. Le stime di Bolin furono di circa $10^{15}$ grammi di carbonio liberati dal biota. Questo tema dominò la conferenza Dahlem, stimolando molte discussioni.

La possibilità che il biota contribuisca significativamente all'aumento nell'atmosfera di anidride carbonica dominò anche due successive conferenze, una allestita dal Department of Energy nel marzo 1977 a Miami Beach e l'altra svoltasi nell'aprile dello stesso anno a Ratzeburg nella Germania Federale organizzata dall'International Council of Scientific Unions. Recenti articoli apparsi su «Science» e su altre pubblicazioni sostengono la conclusione che il biota sia una fonte netta di anidride carbonica atmosferica e non un serbatoio. I modelli degli oceanografi sono chiaramente in discussione. Che cosa significa tutto questo per il bilancio globale del carbonio?

Si è ben lungi dal poter rispondere con chiarezza. Nell'atmosfera entrano quantità di carbonio sostanzialmente maggiori di quante ne vengano accumulate. Oltre ai $5 \times 10^{15}$ grammi di carbonio provenienti annualmente dalla combustione dei combustibili fossili, altri, da $4 \times 10^{15}$ a $8 \times 10^{15}$ grammi, e forse più, possono venire liberati attualmente in conseguenza della distribuzione di foreste e dell'ossidazione accelerata di humus. Di questa quantità combinata variabile da $9 \times 10^{15}$ a $13 \times 10^{15}$ grammi di carbonio per anno, solo $2,3 \times 10^{15}$ grammi si accumulano nell'atmosfera. I restanti $7 \times 10^{15}$ - $11 \times 10^{15}$ grammi, e forse più, vengono immagazzinati da qualche parte sulla Terra. Ma dove? Come abbiamo visto, gli attuali modelli relativi all'assorbimento da parte degli oceani rendono conto dell'allontanamento di meno di $3 \times 10^{15}$ grammi di carbonio per anno. Gli oceanografi stan-

| | MASSA DI VEGETAZIONE ($10^3$ GRAMMI DI CARBONIO PER METRO QUADRATO) |
|---|---|
| FORESTA PLUVIALE EQUATORIALE E TROPICALE | 20 |
| FORESTA PLUVIALE TEMPERATA | 16 |
| FORESTA STAGIONALE, BOSCHI, BOSCAGLIA E SAVANA | 1,8-16 |
| FORESTA SEMPREVERDE | 2-3 |
| FORESTA DECIDUA | 13,5 |
| FORESTA BOREALE E MONTANA | 9 |
| STEPPA E PRATERIA | 0,7 |
| DESERTO ARIDO E SEMIARIDO | <0,3 |
| TUNDRA ARTICA | 0,3 |
| ALTIPIANI CON TUNDRA ALPINA | <0,3 |

**Il carbonio accumulato nelle piante è distribuito come indicato in questa mappa terrestre basata sul lavoro di H. Brockmann Jerosch. La quantità totale di carbonio che si trova racchiuso nella biomassa terrestre è di circa $830 \times 10^{15}$ grammi. In confronto, il carbonio contenuto nella biomassa complessiva degli oceani è trascurabile: meno di $2 \times 10^{15}$ grammi. Circa il 40 per cento di tutto il carbonio contenuto nelle piante è immagazzinato nella foresta pluviale tropicale. Un altro 14 per cento è trattenuto nelle foreste tropicali stagionali. A tutte le latitudini le foreste racchiudono circa il 90 per cento di tutto il carbonio accumulato negli ecosistemi planetari, sia terrestri sia marini.**

**Previsioni circa il consumo di combustibili fossili (*curve in nero*) e il contenuto atmosferico di anidride carbonica (*curve in colore*) in funzione dei tassi di incremento minimi e massimi plausibili. I tassi di consumo dei combustibili sono ricavati da un recente studio effettuato presso l'Oak Ridge National Laboratory. Le proiezioni di «minimo» assegnano un incremento annuo del 2 per cento fino al 2025, seguito da un decremento simmetrico. Le proiezioni di «massimo» assegnano un tasso di crescita del 4,3 per cento per anno fino a che quest'ultima verrà limitata dall'impoverimento di risorse verso la metà del prossimo secolo. L'incertezza viene pesantemente aggravata dalla possibilità, recentemente riconosciuta, che la distruzione di foreste possa anche causare la liberazione di grandi quantità di anidride carbonica nell'atmosfera. La parte della curva a dente di sega rappresenta le misurazioni compiute a Mauna Loa a partire dal 1958. Attualmente, i combustibili fossili che ogni anno vengono bruciati liberano nell'atmosfera una quantità di anidride carbonica equivalente a circa 2 parti per milione. Circa la metà di questa quantità viene trattenuta nell'atmosfera.**

no ora rivedendo le proprie supposizioni per scoprire se hanno trascurato eventuali meccanismi capaci di catturare l'eccedenza di carbonio.

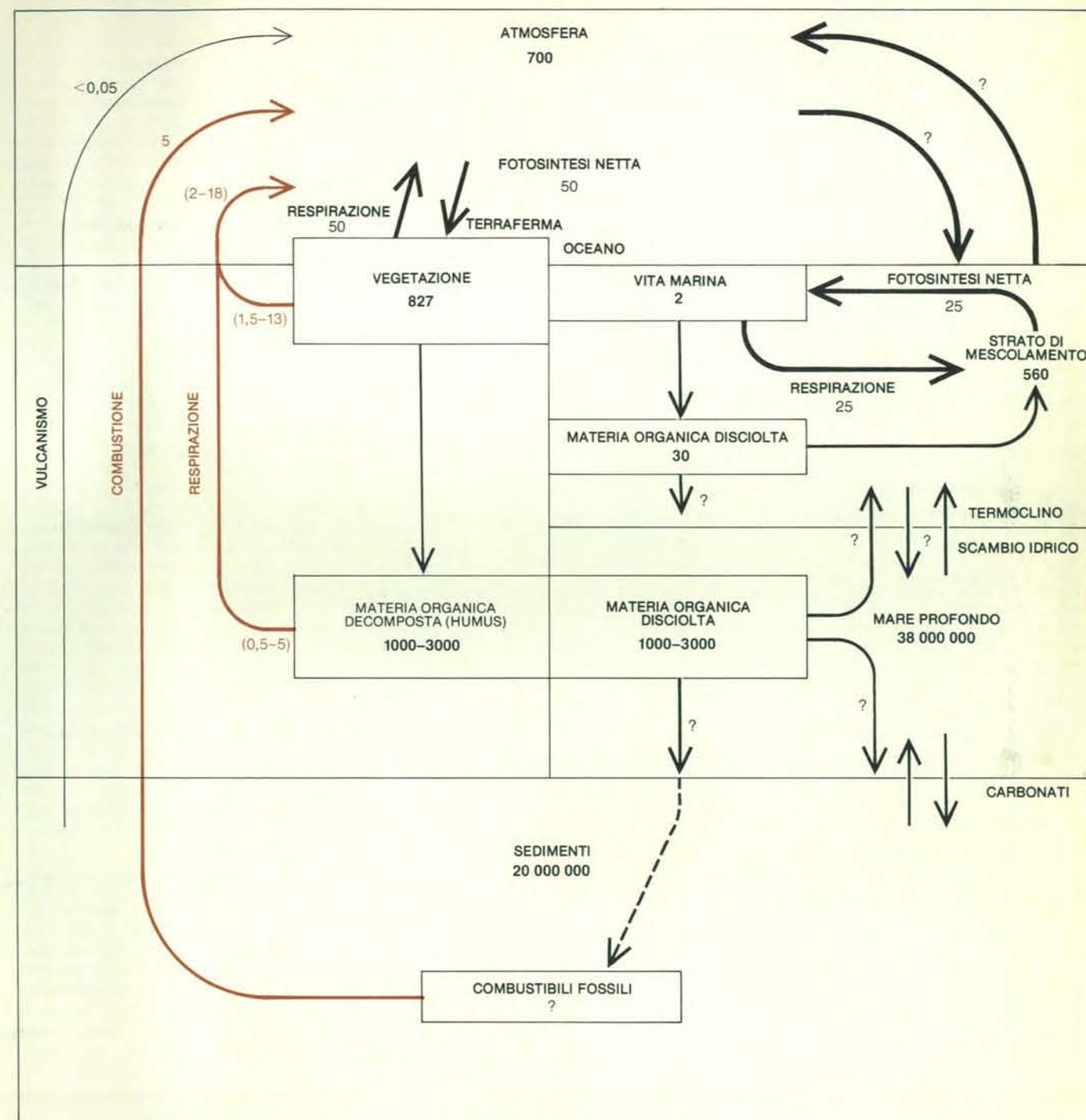
Data l'importanza del problema, coloro che si occupano di ecosistemi terrestri stanno ora attentamente esaminando i propri dati, particolarmente quelli relativi ai cambiamenti nella massa forestale. Come possiamo essere certi che le comunità terrestri siano in realtà una fonte e non un serbatoio? Il problema è stato recentemente affrontato in diversi modi. I dati non sono così attendibili come si potrebbe sperare, ma essi appaiono vincolanti per coloro che li padroneggiano.

L'analisi si basa in primo luogo sulla conoscenza delle grandezze relative dei serbatoi di carbonio in seno al biota e sulla produzione netta di ciascuna delle principali comunità di piante sulla Terra. Per esempio, lo studio di Whittaker e Likens mostra che la foresta pluviale tropicale contiene circa il 42 per cento di tutto il carbonio incorporato nella vegetazione terrestre e contribuisce per il 32 per cento alla produzione primaria netta complessiva. Le aree forestali, di tutti i tipi - tropicale, temperato, boreale - racchiudono il 90 per cento di tutto il carbonio contenuto nella vegetazione e contribuiscono per il 60 per cento alla produzione primaria netta. L'unico altro importante contributore singolo alla produzione primaria netta è la savana più o meno alberata con una quota del 12 per cento, contro soltanto circa il 3 per cento quanto a massa di carbonio incorporata. Tutte le terre coltivate dei continenti concorrono nella misura dell'8 per cento alla produzione primaria netta totale e nella misura di meno dell'1 per cento alla massa di carbonio incorporata. Le stime di Whittaker e Likens si collocano circa a metà strada tra gli estremi dei risultati di altre analisi presentate alla conferenza di Ratzeburg e che furono effettuate sotto la direzione di P. Duvigneaud dell'Università di Bruxelles. Poiché tutti i differenti studi confermano l'importanza delle foreste, particolarmente delle foreste tropicali, è essenziale stabilire se questi ecosistemi stanno cambiando in dimensione e, nel caso affermativo, con quale intensità.

I dati sono scarsi. Esiste l'esperienza di Henry C. Darby, un geografo inglese che nel 1954 pubblicò una valutazione circa i cambiamenti nella vegetazione forestale dell'Europa durante il millennio fra il 900 d.C. e il 1900. Nel corso di questo periodo, la copertura forestale dell'Europa occidentale si era ridotta da circa il 90 per cento a circa il 20 per cento. Una simile modificazione intervenne in epoca precedente nelle regioni del Mediterraneo, particolarmente nel levante. In seguito a questa imponente riduzione dell'area forestale, fu liberata una quantità di carbonio tale da costituire una frazione significativa del carbonio precedentemente contenuto nell'atmosfera. È ragionevole pensare che, a partire dal 1900, la continua industrializzazione e la crescita della popolazione abbiano prodotto simili cambiamenti nelle rimanenti foreste.

Timothy Wood e Daniel B. Botkin, dell'Ecosystems Center a Woods Hole, hanno recentemente compiuto uno studio sui cambiamenti intervenuti nell'area forestale del New England a partire dall'epoca dell'arrivo dei coloni europei. Essi hanno trovato che fino al 1900 si verificò una continua riduzione nella produzione totale fissa delle foreste, ma che, da quel momento, iniziò un periodo di ripristino conseguente all'abbandono dell'agricoltura e alla penetrazione di foreste nei terreni precedentemente coltivati. Tale ripristino, tuttavia, non è stato tale da ristabilire un serbatoio di carbonio equivalente a quello delle foreste originarie. Le foreste sono state regolarmente sfruttate e non hanno mai raggiunto l'altezza e l'estensione di quelle di origine: la produzione fissa di carbonio non va oltre la metà di quella originaria. I dati più recenti indicano che l'allargamento dell'area forestale si è ora arrestata probabilmente a causa di una rinnovata espansione dell'agricoltura e di un intensificato sfruttamento degli alberi.

Lo studio di Wood e Botkin mostra che in una foresta della zona temperata, lasciata in condizione di rigenerarsi da uno sfruttamento intensivo, l'accumulo di carbonio toccava una percentuale del 3 o 4 per cento per anno della produzione primaria netta lungo l'intero periodo del ripristino, circa 70 anni. Se in altre foreste della zona temperata si verificasse un accumulo corrispondente a una frazione simile della produzione primaria netta e se un'analoga quantità addizionale di carbonio si accumulasse nell'humus, la quantità di carbonio accumulatasi in tutte le foreste della zona temperata della Terra raggiungerebbe, annualmente, il valore di $0,5 \times 10^{15}$ grammi. L'esperienza maturata con le foreste del New England suggerisce che la ricrescita di foreste nelle zone temperate probabilmente non costituisce attualmente un grande serbatoio per l'anidride carbonica atmosferica.

Nel frattempo, vi è stata una continua espansione dell'agricoltura in altri territori forestali, un continuo sfruttamento di foreste primarie, un generale inquinamento della Terra in seguito alle attività umane. Le più cospicue irruzioni in ambienti forestali si sono verificate nelle zone tropicali. La più estesa zona ricoperta da foreste che rimane sulla Terra è il bacino amazzonico e per tale motivo siamo andati in cerca di dati sulla velocità di sfruttamento della foresta amazzonica. Non esistono indagini che, con

**Schema del bilancio globale del carbonio che ne mostra i principali «serbatoi» e le velocità di scambio annuale tra i serbatoi che sono in contatto. Le quantità sono espresse in unità di $10^{15}$ grammi, o miliardi di tonnellate. Le immissioni complessive nell'atmosfera controllate dalle attività umane appaiono in colore. Le piante terrestri fissano una quantità netta di circa $50 \times 10^{15}$ grammi di carbonio per anno. Questo carbonio viene sia consumato e immediatamente respirato dai vari organismi terrestri, sia immagazzinato nella massa delle piante. Il bilancio tra fissazione (fotosintesi netta) e immagazzinamento esteso a tutti gli organismi terrestri determina il fatto che vi sia un flusso netto di anidride carbonica verso o dal biota. Molti biologi sono ora del parere che vi sia stato un flusso netto a lungo termine di anidride carbonica dal biota nell'atmosfera e che tale flusso si mantenga. Il carbonio fissato dagli organismi marini viene sia respirato sia immagazzinato. Si ritiene comunemente che la maggior parte di esso venga immediatamente respirata e riciclata. È possibile che l'affondamento delle particelle fecali trascini nelle profondità oceaniche più carbonio di quanto sia stato supposto, il che integrerebbe la normalmente lenta diffusione di anidride carbonica negli strati superficiali degli oceani, dove essa entra in equilibrio con il sistema carbonati-bicarbonati. Anche se le profondità oceaniche costituiscono un serbatoio praticamente illimitato l'anidride deve prima entrare nello strato di mescolamento e poi nel termoclino, strato caratterizzato da stratificazione termica che impedisce il mescolamento con gli strati più profondi.**

sufficiente approssimazione, possano essere applicate all'intero bacino amazzonico senza incorrere in tutti quei rischi che sono connessi con ogni estrapolazione. In ogni caso, Lawrence S. Hamilton della Cornell University e i suoi colleghi hanno recentemente pubblicato una serie di relazioni sulle modificazioni della foresta in territorio venezuelano.

Una delle relazioni, di cui è autore J. P. Veillon, indica che, tra il 1950 e il 1975, si è avuta una riduzione del 33 per cento nella superficie a foresta delle pianure occidentali del Venezuela. Hamilton, in una introduzione alla serie di relazioni, riporta dati della FAO che «suggeriscono un ritmo di disboscamento annuale della foresta umida compreso fra lo 0,6 e l'1,5 per cento della superficie ancora esistente». Studi compiuti su altri settori del bacino amazzonico testimoniano dell'elevato ritmo di sviluppo di autostrade, dell'espansione dell'agricoltura a scapito della foresta e della mancanza di ripristino di vegetazione dopo che la foresta è stata disboscata. Praticamente, nessuno studioso bene informato e competente per quanto riguarda le zone tropicali sostiene che la foresta amazzonica potrà sfuggire a interferenze considerevoli connesse allo sfruttamento di legname e all'apertura di radure durante i prossimi 30 anni.

Le migliori ipotesi di cui si dispone indicano che il tasso di disboscamento delle foreste tropicali per qualunque finalità è probabilmente compreso tra lo 0,5 e l'1,5 per cento della superficie esistente annualmente. Supponendo un tasso dell'1 per cento e ammettendo che la maggior parte del disboscamento sia connessa all'espansione dell'agricoltura e di terreni da pascolo entro zone precedentemente coperte da foreste, la quantità di carbonio che sarebbe resa libera dalla sola riduzione di foreste si aggirerebbe annualmente intorno a $4,5 \times 10^{15}$ grammi. Applicando queste previsioni all'intero biota terrestre e tenendo conto di un accumulo addizionale dovuto a ricrescita, i miei colleghi e io abbiamo recentemente valutato che la liberazione netta di anidride carbonica nell'atmosfera da parte del biota sia di $6 \times 10^{15}$ grammi di carbonio per anno. L'entità di una cessione addizionale in seguito alla decomposizione di humus è difficile da calcolare, tuttavia abbiamo stimato che la perdita sia su livelli di $2 \times 10^{15}$ grammi per anno. La liberazione totale di carbonio è pertanto stimabile in circa $8 \times 10^{15}$ grammi per anno. Le incertezze sono tuttavia così ampie da indurci ad avanzare la supposizione che l'effettiva perdita in seno al biota potrebbe oscillare tra $2 \times 10^{15}$ e $18 \times 10^{15}$ grammi per anno. Anche se occorre tenere sempre presenti i limiti dei dati su cui si basano tali supposizioni, è difficile evitare la conclusione che la distruzione di foreste sulla Terra finisca col provocare un'aggiunta di anidride carbonica nell'atmosfera a una velocità paragonabile a quella con cui lo stesso composto viene liberato per combustione dai combustibili fossili, velocità che assumerebbe valori sensibilmente più elevati qualora venisse compresa l'ossidazione dell'humus.

Queste considerazioni sono state accolte con sorpresa da molti esperti di problemi agricoli e forestali. L'ipotesi di riferimento nella loro esperienza è stata quella secondo cui le moderne pratiche agricole e forestali abbiano incrementato la produttività primaria netta dei sistemi terrestri sotto il controllo umano e che tale incremento debba essersi tradotto in un accumulo di carbonio più veloce nell'humus dei terreni controllati rispetto a quelli naturali. Tuttavia, la produttività media dell'agricoltura misurata secondo le unità degli ecologi, in termini di materia organica totale disponibile per l'uomo o per gli altri animali, è sostanzialmente inferiore alla produttività primaria media delle comunità naturali soppiantate dall'agricoltura. Le comunità di piante agricole non sono sfruttate col fine di rendere rapido l'avvicendamento di serbatoi di carbonio durante la loro utilizzazione da parte dell'uomo. Ciò significa che i terreni adibiti a uso agricolo non accumulano tanto carbonio quanto le foreste che hanno rimpiazzato. Analogamente, le praterie convertite all'agricoltura tendono a perdere i materiali organici del terreno per decomposizione e non ad accumularne di addizionali.

Le foreste sfruttate in modo intensivo, anche se possono produrre una quantità maggiore di legno, tendono ad avere, al tempo della raccolta del legname, un rendimento fisso molto inferiore a quello delle foreste primarie di cui hanno preso il posto. Ciò nuovamente significa che il tempo di avvicendamento è stato ridotto e che la «dotazione» fissa di carbonio non raggiunge mai la stessa entità raggiunta nelle foreste originarie. Pertanto, la sostituzione di foreste primarie con foreste secondarie sfruttate come fonte di legname e di pasta di legno ha come risultato una liberazione netta di anidride carbonica. Un'altra fonte netta di anidride carbonica è la decomposizione dell'humus. La decomposizione viene stimolata dall'intervento di raccolta del legname e può protrarsi parecchi anni dopo tale raccolta, durante i primi stadi della successione di specie che seguono l'abbattimento. Tutti questi fattori producono una fuoriuscita netta di carbonio in tutti i casi in cui una foresta vergine primaria viene rimpiazzata da una foresta secondaria sfruttata dall'uomo.

Un'altra circostanza sembra dare sostegno alla conclusione che il biota terrestre è stato per molti decenni una fonte netta di anidride carbonica piuttosto che un serbatoio di accumulo. Confrontando i rapporti tra gli isotopi del carbonio in alcuni alberi, Minze Stuiver, dell'Università di Washington, ha calcolato che nel secolo compreso tra il 1850 e il 1950 il biota sembra aver lasciati liberi nell'atmosfera $1,2 \times 10^{15}$ grammi di carbonio per anno. Nello stesso periodo, le immissioni connesse ai combustibili fossili sono state in media di $6 \times 10^{15}$ grammi di carbonio per anno.

Stuiver sfruttò la circostanza per cui il rapporto tra il carbonio 12 e il carbonio 13 varia tra l'atmosfera, il biota e i combustibili fossili. Il biota e i combustibili fossili sono leggermente più ricchi dell'isotopo più leggero, carbonio 12. Inoltre, si verifica che un terzo isotopo del carbonio, il carbonio 14, si produce negli strati superiori dell'atmosfera in seguito al bombardamento da parte dei raggi cosmici dell'isotopo comune dell'azoto, azoto 14. Esso è stato anche prodotto in grandi quantità nel corso degli esperimenti nucleari nell'atmosfera. Poiché il carbonio 14 ha una emivita di circa 6000 anni, è scomparso da lungo tempo dai combustibili fossili, la cui epoca di formazione risale a milioni di anni fà. Dalla combustione dei combustibili fossili, si libera una «partita» di carbonio deficiente di carbonio 14, che tende a diluire il carbonio 14 nell'atmosfera. Misurando la concentrazione di differenti isotopi del carbonio negli anelli di alberi di età conosciuta e confrontandone i rapporti con quelli relativi all'atmosfera e ai combustibili fossili, Stuiver è stato in grado di stimare la quantità di carbonio fuoriuscito dal biota. Le misure sono tecnicamente difficili da eseguire; inoltre, i rapporti tra gli isotopi del carbonio negli alberi sono influenzati da numerosi fattori ambientali che rendono i risultati meno netti di quanto si vorrebbe. Ciononostante, questa tecnica fornisce un importante metodo addizionale per valutare l'entità della fuoriuscita di anidride carbonica dal biota.

Alla luce dell'evidenza, difficilmente si può mettere in discussione il fatto che l'aumento del tenore atmosferico di anidride carbonica dipende non solo dalla combustione di combustibili fossili, ma anche dalla distruzione di foreste. G. Evelyn Hutchinson, della Yale University, fece questa considerazione in un capitolo di *The Earth as a Planet* un libro pubbblicato da Gerard P. Kuiper nel 1954. Hutchinson fece la supposizione che la quantità di carbonio resa libera dal biota fosse circa uguale a quella emessa in seguito alla combustione di combustibili fossili. Le indicazioni più attendibili che possediamo suggeriscono che questa relazione sussiste.

C'è abbastanza carbonio trattenuto nel biota perché tale relazione si mantenga per un altro decennio o due, in concomitanza col previsto picco nei consumi mondiali di petrolio. Per quell'epoca, se non prima, la Terra sarà stata sospinta verso cambiamenti climatici come conseguenza dell'accumularsi di anidride carbonica nell'atmosfera - a patto che tali cambiamenti si verifichino realmente. Il fatto che ciò accada, oppure no, dipenderà dalla scala dell'effetto che produrrà l'anidride carbonica, una stima che è estremamente difficile da formulare con qualche sicurezza. Ora si sa che il clima è influenzato da cambiamenti nell'emissione solare, da cambiamenti nella riflettività terrestre e da altri fattori. Rimane da verificare se il contenuto atmosferico di anidride carbonica sarà abbastanza elevato da porsi come fattore determinante. Se l'effetto dell'anidride carbonica sarà veramente dominante, è probabile che la Terra si riscalderà in modo differenziale, con aumenti di temperatura crescenti verso i poli. È presumibile che un simile cambiamento possa sospingere le zone desertiche verso i poli, ampliando le regioni aride e restringendo le regioni adatte all'agricoltura. Tale prospettiva non è certo incoraggiante per un mondo la cui popolazione umana potrà forse raddoppiare nei prossimi 30-35 anni.

Se si fosse sicuri dell'eventualità di un cambiamento climatico preoccupante nei prossimi decenni il corso da impartire alle azioni sarebbe abbastanza chiaro. La combustione di combustibili fossili verrebbe contenuta, così da limitare questa fonte di anidride carbonica. Energiche iniziative verrebbero inoltre intraprese per impedire lo sfruttamento di foreste primarie in tutto il mondo, per ampliare le aree destinate a foreste e per permettere a queste aree di sviluppare alberi di elevato rendimento fisso. È assai dubbio se queste drastiche misure potrebbero venire messe in pratica; i problemi sociali che comporterebbero sarebbero evidentemente profondi.

Altri suggerimenti sono stati avanzati, compreso il seguente: poiché la disponibilità di fosforo è da qualcuno considerata fattore limitante della produzione primaria netta negli oceani, le nazioni avanzate dovrebbero destinare una parte della propria energia industriale all'estrazione del fosforo, con lo scopo di trasferirlo il più rapidamente possibile agli oceani relativamente infecondi, così, da stimolare l'attività fotosintetica e l'accumulo di carbonio. Questo progetto è superficialmente attraente, in quanto sembra fornire un mezzo per accelerare l'accumulo di carbonio nel mare. Esiste tuttavia la possibilità che ogni incremento di fotosintesi indotto negli oceani venga annullato da un equivalente intensificazione della respirazione, per cui non si verificherebbe alcun arricchimento netto di carbonio. Esistono molti interrogativi circa la realizzabilità di questo intervento, compreso quello, fondamentale, che nasce dalla domanda se il fosforo sia veramente un fattore limitante critico nelle acque degli oceani. Esiste una possibilità ragionevole che anche l'azoto sia un fattore limitante, nel qual caso la stimolazione della fotosintesi nel mare potrebbe rivelarsi sostanzialmente più difficile di quanto supposto.

Tuttavia, recenti esperienze hanno sottolineato all'attenzione di tutti noi che lavoriamo in questi campi il fatto che attualmente rimangono ancora in gran parte sconosciuti importanti aspetti del bilancio terrestre del carbonio. Si potrebbe progredire nella conoscenza di questi aspetti attraverso un rilevamento dei cambiamenti nelle zone di diffusione e nella struttura delle foreste in tutta la Terra per mezzo di fotografie riprese da satelliti. Si stanno anche compiendo indagini per stabilire se certi meccanismi biotici potrebbero facilitare il trasferimento di carbonio nelle profondità oceaniche in quantità considerevolmente maggiori di quanto supponiamo avvenga attualmente, ma la prospettiva immediata di conoscere tutti gli aspetti del bilancio terrestre del carbonio non è brillante.

I rischi potenziali connessi a un aumento costante nel contenuto atmosferico di anidride carbonica si manifesteranno gravi nei prossimi decenni e imporranno senza dubbio pesanti ipoteche su decisioni quali quella se si debba intensificare lo sviluppo di centrali nucleari invece di quelle a combustibili fossili e se si debbono conservare le aree forestali invece di esercitarvi interferenze. Quasi non esistono aspetti di politica nazionale e internazionale che possano non rimanere coinvolti dalla prospettiva di una modificazione climatica globale. Può darsi che l'anidride carbonica, finora considerata un gas innocuo presente in tracce nell'atmosfera, vada a collocarsi in una posizione cruciale di minaccia nei confronti dell'attuale ordine mondiale.

| | AREA ($10^6$ CHILOMETRI QUADRATI) | PRODUZIONE PRIMARIA NETTA ($10^{15}$ GRAMMI DI CARBONIO PER ANNO) | MASSA VEGETALE ($10^{15}$ GRAMMI DI CARBONIO) |
|---|---|---|---|
| FORESTA PLUVIALE TROPICALE | 17,0 | 16,8 | 344,0 |
| FORESTA STAGIONALE TROPICALE | 7,5 | 5,4 | 117,0 |
| FORESTA TEMPERATA SEMPREVERDE | 5,0 | 2,9 | 79,0 |
| FORESTA TEMPERATA DECIDUA | 7,0 | 3,8 | 95,0 |
| FORESTA BOREALE | 12,0 | 4,3 | 108,0 |
| BOSCHI E BOSCAGLIA | 8,5 | 2,7 | 22,0 |
| SAVANA | 15,0 | 6,1 | 27,0 |
| PRATERIA TEMPERATA | 9,0 | 2,4 | 6,3 |
| TUNDRA E PASCOLI ALPINI | 8,0 | 0,5 | 2,3 |
| VEGETAZIONE DESERTICA | 18,0 | 0,7 | 5,9 |
| ROCCE, GHIACCIO E SABBIA | 24,0 | 0,03 | 0,2 |
| TERRA COLTIVATA | 14,0 | 4,1 | 6,3 |
| PALUDI E ACQUITRINI | 2,0 | 2,7 | 13,5 |
| LAGHI E CORSI D'ACQUA | 2,0 | 0,4 | 0,02 |
| TOTALE CONTINENTALE | 149,0 | 52,8 | 826,5 |
| MARE APERTO | 332,0 | 18,7 | 0,45 |
| ZONE DI RISALITA | 0,4 | 0,1 | 0,004 |
| PIATTAFORMA CONTINENTALE | 26,6 | 4,3 | 0,12 |
| LETTO DI ALGHE E SCOGLIERA | 0,6 | 0,7 | 0,54 |
| ESTUARI | 1,4 | 1,0 | 0,63 |
| TOTALE OCEANICO | 361,0 | 24,8 | 1,74 |
| TOTALE MONDIALE | 510,0 | 77,6 | 828,0 |

**Le principali comunità di piante della Terra sono classificate secondo la loro estensione, la loro produzione primaria netta e la quantità di carbonio che incorporano. La produzione primaria è la quantità di carbonio che una comunità di piante rende annualmente disponibile per essere raccolta o per fungere da nutrimento a numerosi organismi consumatori, sia selvatici che domestici. Anche se soltanto il 30 per cento circa della Terra è costituito da terre emerse, la produzione primaria netta della vegetazione continentale è leggermente più che doppia rispetto alla produzione primaria della vegetazione marina. La quantità di carbonio accumulata nelle piante terrestri è circa 500 volte maggiore di quella accumulata negli ecosistemi marini ed è approssimativamente uguale alla quantità di carbonio presente nell'atmosfera. I dati in tabella sono stati recentemente compilati da R.H. Whittaker e Gene F. Likens della Cornell University.**

# Come conservare l'energia

***È possibile disporre di quantità di energia adeguate per lo sviluppo economico senza accrescere in proporzione il consumo delle risorse energetiche non rinnovabili; questo consumo potrebbe anzi diminuire***

di Gian Battista Zorzoli

Uno dei concetti base impartiti a uno studente di materie tecnico-scientifiche è quello di «rendimento di una trasformazione energetica», più comunemente chiamato, nella sua versione ellittica, «rendimento» *tout-court*. Apparentemente si tratta di una definizione univoca, secondo la quale il rendimento $\eta$ è espresso dal rapporto tra l'energia disponibile dopo la trasformazione e l'energia disponibile prima della trasformazione.

In quest'accezione, di rendimento si parla nel linguaggio tecnico: così il rendimento di una caldaia è dato dal rapporto fra l'energia termica trasferita al fluido termovettore e l'energia chimica contenuta nel combustibile utilizzato dalla caldaia stessa. Una volta prescelto il sistema per la trasformazione energetica (nel caso specifico una caldaia), il rendimento misura quindi le prestazioni del sistema stesso in termini di conservazione dell'energia.

Esso sarà di norma un numero minore di uno, in quanto parte dell'energia contenuta nel combustibile andrà «persa» per combustione imperfetta, per trasmissione di calore all'esterno attraverso le pareti della caldaia, per il calore non recuperato contenuto nei gas di scarico. In un caso ideale, si potrà quindi realizzare un valore di rendimento $\eta = 1$. Ciò potrebbe verificarsi, per esempio, in un motore elettrico ideale, privo di perdite, in cui tutta l'energia elettrica in ingresso si trasformi in energia meccanica; tutta l'energia, cioè, si conserverebbe all'interno del sistema.

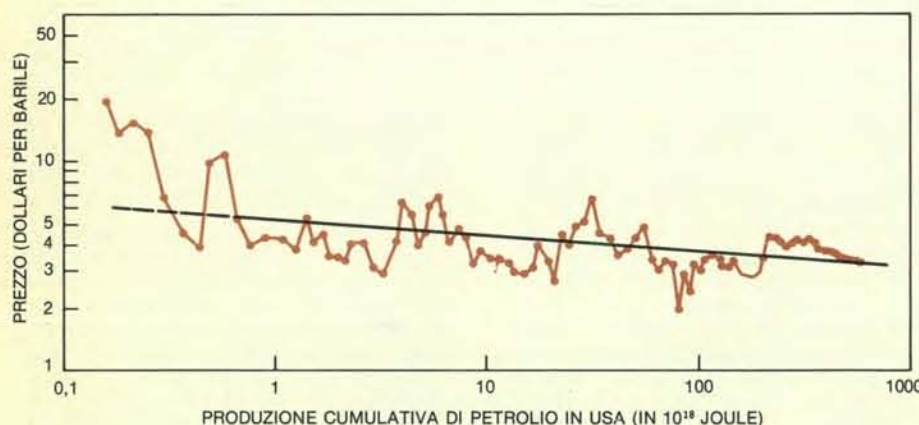
Affinché lo scambio termico si verifichi, la temperatura del fluido secondario in uscita da uno scambiatore di calore deve però essere inferiore alla temperatura di ingresso del fluido primario. Di conseguenza l'energia disponibile per compiere un lavoro utile è minore utilizzando il fluido secondario rispetto al caso in cui per compiere tale lavoro si sfrutti il fluido primario. Il valore del rendimento della trasformazione energetica, secondo la definizione sopra data, non fornisce alcuna informazione su questa perdita di «qualità» dal punto di vista dell'utilizzazione dell'energia in gioco. Se addirittura si modificassero le caratteristiche del fluido secondario, per esempio aumentandone la portata, la corrispondente diminuzione della temperatura in uscita e del lavoro utile realizzabile non verrebbe in alcun modo segnalata dal rendimento che - così come lo abbiamo definito - rimarrebbe rigorosamente uguale a uno. La ragione di tale insensibilità risiede nel fatto che la definizione suddetta discende dal primo principio della termodinamica, il quale sancisce la conservazione dell'energia in ogni processo di trasformazione.

Se viceversa facciamo riferimento al secondo principio della termodinamica, secondo il quale è impossibile realizzare una trasformazione energetica in cui tutto il calore prelevato da una sorgente calda sia convertito in lavoro, e ci preoccupiamo di raggiungere da questo punto di vista il rendimento massimo, allora quest'ultimo sarà definito dal rapporto $\varepsilon$ tra l'energia (o lavoro) utile convertita (o convertito) dal sistema in esame e la massima energia (o lavoro) utile convertibile, dove evidentemente il secondo termine si riferisce a quello fra i possibili sistemi in grado di realizzare un determinato compito, che massimizza la conversione medesima con lo stesso tipo di energia in ingresso. Con questa definizione di rendimento, consideriamo i modi alternativi di utilizzazione della tipica fonte energetica primaria (un combustibile fossile). Partendo sempre da un processo di combustione che avviene a temperatura relativamente elevata (intorno ai 1000 °C) si può produrre il vapore surriscaldato necessario per la generazione di energia elettrica con un buon rendimento $\eta$, dopo di che il calore residuo a relativamente bassa temperatura (più della metà dell'energia disponibile nella fonte primaria) potrebbe essere per esempio usato per il riscaldamento di edifici; oppure si può produrre direttamente acqua calda intorno agli 80 °C per il puro e semplice riscaldamento di edifici. In questo caso, anche se la caldaia avesse un rendimento $\eta = 0{,}75$-$0{,}80$, il processo di trasformazione energetica misurato sulla base del secondo principio della termodinamica avrebbe un rendimento $\varepsilon$ intorno a 0,1.

Tradotto quindi in termini pratici, il diverso valore di $\varepsilon$ sta a indicare il fatto che, poiché l'energia di un fluido in sé non ha valore, questo derivandogli dalla capacità di compiere lavoro, cioè dallo scarto fra la temperatura a cui il fluido si trova e la temperatura ambiente, in linea di principio è poco conveniente utilizzare un combustibile fossile (in grado di fornire energia ad alta temperatura) per avere a disposizione energia a temperatura relativamente bassa. Tuttavia queste considerazioni, abbastanza ovvie, sono state sino a oggi sostanzialmente ignorate nelle applicazioni pratiche (e anche sui manuali di fisica, dove non manca mai la definizione di $\eta$, mentre è assente quella di $\varepsilon$). La ragione dell'interesse esclusivo per il rendimento derivato dal primo principio della termodinamica è abbastanza evidente. Una utilizzazione di una certa fonte di energia adeguata al suo pregio interessa soltanto se essa è non rinnovabile e relativamente scarsa. Poiché l'utente misura la «scarsità» di una risorsa attraverso il parametro economico del suo prezzo di mercato, la tendenza storica verso costi reali decrescenti delle fonti energetiche non rinnovabili (valida sino all'inizio degli anni settanta ed esemplificata dalla figura in basso nella pagina a fronte) è paradossalmente equivalsa a un accrescimento apparente delle risorse residue man mano che tali risorse venivano consumate. Donde il disinteresse teorico e pratico per il rendimento inteso come $\varepsilon$ e l'identificazione univoca del rendimento in $\eta$, che, misurando la capacità di un sistema di trasformare in modo ottimale l'energia fornita in ingresso, rappresenta un dato essenziale nella determinazione delle caratteristiche del sistema in esame atte a minimizzarne il costo.

Una conferma di quest'interpretazione viene dal fatto che l'interesse scientifico (e tecnico-pratico) per una valutazione dei processi di trasformazione di energia basata su $\varepsilon$ si è sviluppato solo con l'insorgere della crisi energetica, come conferma in particolare lo studio su questo tema promosso dalla American Physical Society in seguito a un incontro scientifico svoltosi a Los Alamos nell'agosto 1973. Tra l'altro l'utilizzazione del rendimento $\varepsilon$ presenta il vantaggio di eliminare l'anomalia in cui si incorre quando si definiscono le prestazioni di una pompa di calore sulla base del primo principio della termodinamica; in tal caso, invece di fare ricorso a definizioni come quella di «coefficiente di amplificazione» (*si veda l'illustrazione qui in basso*) per indicare lo sfruttamento di lavoro meccanico al fine di trasferire energia termica da bassa ad alta temperatura, si verifica che anche la pompa di calore, inserita in un sistema atto a realizzare un certo compito, contribuirà a un rendimento complessivo $\varepsilon$ sicuramente infe-

| SISTEMI | RENDIMENTO ($\eta$) | RENDIMENTO ($\varepsilon$) |
|---|---|---|
| SCALDABAGNO ELETTRICO | 0,75 | 0,015 |
| SCALDABAGNO A GAS | 0,50 | 0,029 |
| RISCALDAMENTO CENTRALIZZATO | 0,75 | 0,145 |
| CALORE DI PROCESSO INDUSTRIALE | 0,75 | 0,250 |

**In questa tabella sono riportati alcuni valori di rendimenti di sistemi. Tali valori sono stati calcolati sulla base del primo ($\eta$) e del secondo principio della termodinamica ($\varepsilon$). È interessante notare che, a parità di rendimento $\eta$, il rendimento $\varepsilon$ può variare addirittura di un fattore 17 (da 0,015 a 0,25).**

**Con l'aumento della quantità di petrolio estratto dai pozzi americani il suo prezzo è andato diminuendo, come indica il grafico i cui estremi si riferiscono, a sinistra, al 1869 e, a destra, al 1971.**

| SETTORE | DOMANDA IN MILIONI DI TONNELLATE DI PETROLIO EQUIVALENTE | | VALORE PERCENTUALE | |
|---|---|---|---|---|
| INDUSTRIA (NON DELL'ENERGIA) | 320 | | 34,9 | |
| INDUSTRIA DELL'ENERGIA* | 168 | | 18,2 | |
| TRASPORTI | 120 | | 13,1 | |
| — SU STRADA | | — 94 | | — 10,3 |
| — ALTRI | | — 26 | | — 2,8 |
| ALTRI SETTORI** | 308 | | 33,7 | |
| — RISCALDAMENTO E CONDIZIONAMENTO | | — 216 | | — 23,6 |
| — ALTRO | | — 92 | | — 10,1 |
| TOTALE | 916 | | 100% | |

* INCLUDE TUTTE LE PERDITE DI ENERGIA PER TRASFORMAZIONE E DISTRIBUZIONE.
** RESIDENZIALE, COMMERCIALE, ARTIGIANALE, PUBBLICA AMMINISTRAZIONE.

**Quasi un quinto della domanda all'interno della Comunità Economica Europea è assorbita dall'industria dell'energia: raffinerie, centrali elettriche, linee di trasporto e di distribuzione dell'energia elettrica, ecc. In altri termini, la conversione dell'energia primaria nelle forme finali in cui viene utilizzata «distrugge» circa un quinto dell'energia inizialmente disponibile. Si tratta quindi di un settore produttivo in cui esistono larghi margini per contribuire a una migliore utilizzazione delle fonti primarie di energia, in particolare se si sfrutta il calore residuo della generazione di energia elettrica, pari a due volte quest'ultima nel caso di centrali nucleari e a una volta e mezza nel caso di centrali termiche convenzionali. I dati si riferiscono al 1974.**

**Nella pompa di calore si effettua il trasferimento di energia termica da una temperatura relativamente bassa (per esempio 10 °C) a una temperatura più elevata (per esempio 60 °C). Ciò avviene utilizzando come intermediario un fluido come l'ammoniaca capace di evaporare a temperature piuttosto basse anche a pressioni vicine all'atmosferica. Un simile fluido viene appunto immesso in uno scambiatore di calore, in cui assorbe energia termica, da un ambiente esterno che si trova a bassa temperatura, evapora e, successivamente, viene portato a pressione più elevata in un apposito compressore. Poiché in quest'ultimo processo il fluido aumenta anche di temperatura, esso può cedere energia termica a un altro ambiente esterno, riscaldandolo, dopo di che viene fatto espandere e il ciclo ricomincia. L'efficienza della pompa di calore è misurata dal «coefficiente di amplificazione», sempre superiore all'unità, in quanto l'energia termica ceduta nel condensatore proviene non solo dall'energia termica fornita dall'ambiente esterno a bassa temperatura, ma anche dall'energia meccanica trasferita dal compressore al fluido.**

riore a uno. Va infine notato come, sulla base della interpretazione statistica dei concetti di lavoro e calore e delle analogie fra meccanica statistica e teoria delle informazioni, si può affermare che l'entropia di un sistema diminuisce di una quantità pari all'informazione che si ha sullo stato del sistema. Le fonti di energia non rinnovabili, poiché si trovano nello stato entropico minimo rispetto alle successive trasformazioni energetiche che da esse possono essere avviate, si presentano pertanto con un elevato contenuto di informazione.

La massimizzazione di $\varepsilon$, cioè la più efficace utilizzazione di una determinata fonte di energia per un particolare compito, può essere allora definita come quella trasformazione corrispondente alla richiesta da parte dell'utenza di un contenuto di informazione il più possibile vicino al contenuto disponibile nella fonte stessa, così da disperdere la quantità minima di informazione. Mediante una utilizzazione congiunta di entrambe le definizioni di $\varepsilon$ e della definizione di $\eta$, è possibile affrontare in modo coerente il problema della gestione ottimale delle risorse energetiche. Questo naturalmente, con l'assunzione implicita che le indicazioni così acquisite andranno verificate sotto il profilo della loro convenienza economica.

SENZA FRANGISOLE

CON FRANGISOLE

**La diversa altezza del sole, più bassa sull'orizzonte d'inverno che d'estate, fa sì che una parete verticale esposta a sud riceverà d'inverno un flusso energetico piuttosto elevato. Se tale parete è dotata di un'ampia vetrata, quest'ultima consentirà l'ingresso della radiazione solare, che - per il cosiddetto «effetto serra» - resterà in larga misura intrappolata all'interno della stanza. Viceversa la sporgenza frangisole, realizzabile per esempio con un balcone, permetterà di sfruttare la maggiore altezza sull'orizzonte del sole durante il periodo estivo per schermare la stanza, in modo da garantire un minimo di condizionamento naturale.**

La gestione ottimale delle risorse energetiche può venire realizzata operando a tre diversi livelli: *a*) intervenendo sull'uso di un sistema energetico dato, inserito in un'organizzazione complessiva delle trasformazioni energetiche, pure data; *b*) intervenendo sullo specifico sistema energetico senza toccare l'organizzazione complessiva delle trasformazioni energetiche in cui esso è inserito (come sottoprodotto può derivare una modifica nell'uso del sistema stesso); *c*) intervenendo sull'organizzazione complessiva delle trasformazioni energetiche, con le possibili conseguenti modifiche dei sistemi in cui essa si articola e dei relativi usi.

Il livello *a* è chiaramente il più elementare, ma con effetti immediati, purché gli utenti di uno specifico sistema intendano utilizzarlo in modo razionale dal punto di vista energetico. Tipici esempi sono l'invito a chiudere gli scuri delle finestre nelle ore notturne per ridurre le dispersioni termiche verso l'esterno degli edifici, a intensificare la manutenzione dell'impianto di riscaldamento, a guidare l'auto su strada aperta intorno alla velocità economica e in città evitando partenze o frenate brusche. Si tratta in sostanza di adoperarsi affinché il sistema dato funzioni il più possibile vicino al rendimento massimo $\eta$ per cui è stato progettato: qualsiasi scostamento dalle condizioni ottimali di funzionamento rappresenta pertanto uno spreco, determinato dal comportamento soggettivo dell'utente. Si è scelta la parola comportamento e non volontà, per sottolineare l'esistenza di condizionamenti esterni al modo di agire dell'individuo: per esempio la pubblicità automobilistica, così ricca di riferimenti alla capacità di ripresa dei motori, non è fatta certo per convincere gli acquirenti sull'opportunità di una guida regolare. Di qui i limiti non solo quantitativi, ma anche di garanzia nella risposta, che si aggiungono alla già ridotta portata teorica di questi interventi.

Di maggior rilievo è indubbiamente il livello *b*, anche se ovviamente sono necessari tempi tecnici (e non solo tecnici)

| SETTORI PRODUTTIVI | ADDETTI OCCUPATI/ VALORE AGGIUNTO | CONSUMI ENERGETICI TOTALI/ VALORE AGGIUNTO | INVESTIMENTI/ ADDETTI OCCUPATI | CONSUMI ENERGETICI TOTALI/ ADDETTI OCCUPATI |
|---|---|---|---|---|
| INDUSTRIE ALIMENTARI | 0,17 | 1,54 | 4,41 | 10,7 |
| MOLITORIA E PASTARIA | 0,17 | 1,83 | 3,99 | 10,4 |
| DOLCIARIA | 0,20 | 0,66 | 2,98 | 4,5 |
| CONSERVIERA | 0,20 | 1,33 | 2,89 | 6,7 |
| CASEARIA | 0,19 | 1,74 | 2,87 | 8,9 |
| GRASSI ANIMALI E VEGETALI | 0,13 | 2,90 | 4,69 | 22,8 |
| ZUCCHERIERA E VARIE | 0,13 | 5,20 | 9,77 | 39,7 |
| BEVANDE ALCOLICHE E ANALCOLICHE | 0,12 | 0,92 | 6,83 | 7,8 |
| TABACCO | 0,27 | 0,40 | 1,65 | 1,5 |
| INDUSTRIE TESSILI | 0,25 | 1,43 | 2,16 | 5,2 |
| SERICA | 0,19 | 2,79 | 3,63 | 14,8 |
| COTONIERA | 0,25 | 2,08 | 2,19 | 8,3 |
| FIBRE CHIMICHE | 0,21 | 0,91 | 2,70 | 4,4 |
| LANIERA | 0,23 | 2,08 | 2,27 | 6,8 |
| CANAPA, LINO, JUTA | 0,25 | 1,58 | 1,98 | 2,6 |
| VARIE | 0,29 | 1,08 | 1,59 | 1,6 |
| VESTIARIO E ABBIGLIAMENTO | 0,35 | 0,28 | 0,71 | 0,8 |
| INDUSTRIE METALLURGICHE | 0,13 | 7,29 | 11,15 | 55,1 |
| INDUSTRIE MECCANICHE | 0,20 | 0,64 | 2,25 | 3,2 |
| MACCHINE ELETTRICHE | 0,21 | 0,32 | 1,85 | 1,6 |
| MACCHINE NON ELETTRICHE | 0,20 | 0,80 | 2,38 | 4,0 |
| MECCANICA DI PRECISIONE | 0,18 | 0,18 | 2,26 | 1,0 |
| VARIE | 0,23 | 1,99 | 3,74 | 8,8 |
| COSTRUZIONE MEZZI DI TRASPORTO | 0,21 | 0,77 | 3,73 | 3,7 |
| LAVORAZIONE MINERALI NON METALLICI | 0,18 | 4,56 | 3,97 | 25,2 |
| INDUSTRIE CHIMICHE | 0,12 | 10,24 | 8,60 | 84,8 |
| DERIVATI DEL PETROLIO E DEL CARBONE | 0,06 | 43,08 | 34,79 | 658,8 |
| CELLULOSA | 0,19 | 3,38 | 8,06 | 17,6 |
| INDUSTRIA CARTARIA | 0,15 | 5,47 | 4,43 | 35,7 |
| INDUSTRIA POLIGRAFICA ED EDITORIALE | 0,15 | 0,24 | 2,02 | 1,5 |
| INDUSTRIA CALZATURIERA | 0,37 | 1,81 | 0,62 | 4,9 |
| PELLI E CUOIO | 0,26 | 0,87 | 1,36 | 3,4 |
| MOBILI E ARREDAMENTI LEGNO | 0,27 | 0,56 | 1,78 | 2,0 |
| LEGNO | 0,25 | 0,60 | 2,39 | 2,5 |
| GOMMA | 0,20 | 0,77 | 4,56 | 3,8 |
| INDUSTRIE FOTO-FONOGRAFICHE | 0,14 | 0,18 | 2,12 | 1,3 |
| MATERIE PLASTICHE VARIE | 0,22 | 1,06 | 2,85 | 4,8 |
| INDUSTRIA MANIFATTURIERA | 0,19 | 3,43 | 3,78 | 17,8 |

**L'insieme dell'industria manifatturiera italiana per milione di lire di valore aggiunto occupa 0,19 persone e utilizza in un anno 3,43 tonnellate di petrolio equivalente. Se confrontata con quella di altri paesi industrializzati la situazione italiana risulta quindi caratterizzata da un elevato fabbisogno di energia per produrre una unità di valore aggiunto. Questo dato non riflette principalmente un ritardo tecnologico quanto scelte produttive squilibrate. Lo sviluppo industriale italiano è stato infatti in larga misura concentrato in settori quali la raffinazione del petrolio, la chimica primaria, la siderurgia, cioè proprio in quelli caratterizzati da più elevata domanda di energia per unità di valore aggiunto. Nella tabella gli addetti e i consumi sono indicati in unità-consumi tonnellate di petrolio equivalente; altri dati in milioni di lire del 1973.**

perché ne vengano avvertiti gli effetti sulla gestione delle risorse energetiche. Per valutare i settori in cui è più conveniente intervenire, è opportuno conoscere la distribuzione della domanda di energia, che è sintetizzata, nel caso della Comunità Economica Europea, nella tabella a pagina 21.

Anche se il settore più rilevante in termini percentuali appare quello dell'industria non dell'energia, si tratta pur tuttavia di un insieme molto diversificato, che richiede interventi specifici per sottosettori specifici. Di conseguenza il settore omogeneo di maggior peso sulla domanda di energia (poco meno di un quarto del totale) appare quello del riscaldamento di edifici adibiti ad abitazione, ad attività artigianale o commerciale e a ospitare strutture della pubblica amministrazione (il condizionamento ha infatti in Europa un'incidenza limitata).

Sulla base delle tipologie edilizie e delle caratteristiche climatiche presenti nei paesi della CEE, si ritiene che circa il 60 per cento della trasmissione di energia termica all'esterno avvenga attraverso i muri periferici, il tetto, il pavimento, mentre il restante 40 per cento è quasi equamente ripartito tra la ventilazione e i vetri delle finestre. Quanto poco si sia fatto per limitare questo fenomeno, lo conferma il fatto che le perdite di energia superano in media il 50 per cento dell'energia impiegata, a causa principalmente dell'insufficiente isolamento termico degli edifici, a cui si aggiunge come causa complementare il modo in cui è progettato ed esercitato il sistema di riscaldamento. Partendo da quest'ultimo problema, si possono considerare le seguenti cause principali di prestazione energetica non ottimale: *a*) le temperature ambientali, sovente mantenute a valori eccessivi; la diminuzione media di 1 °C può consentire di ridurre il consumo di combustibile di circa il 10 per cento; *b*) la scarsa manutenzione dell'impianto, a cui già si è fatto cenno; indagini parziali condotte recentemente in città italiane hanno messo in evidenza rendimenti effettivi di grosse caldaie per riscaldamento intorno a 0,45-0,50, contro valori di targa fra 0,70 e 0,80; *c*) il dimensionamento delle caldaie, orientato al fine di renderle capaci di fornire il calore richiesto per condizioni climatiche abbastanza eccezionali (che negli Stati Uniti, per esempio, vengono annualmente raggiunte o superate per sole 50 ore circa!); *d*) il mancato recupero dell'energia termica dai fumi al camino; *e*) i modi con cui avviene la regolazione dell'impianto termico.

Sui punti *a* e *b* si può agire anche per impianti esistenti; uno stimolo a ridurre la temperatura nei locali durante la loro occupazione e ad abbassarla ulteriormente in caso di prolungata assenza può venire per esempio dall'installazione di contatori termici per ogni singolo appartamento. Anche per il punto *e* è possibile concepire interventi in edifici già costruiti, ma la forma più sofisticata (e costosa) di controllo - nota come termoregolazione - può non fornire vantaggio alcuno, anzi, peggiorare la situazione, se installata in un edificio non accuratamente progettato dal punto di vista termico (qualora per esempio il sottotetto sia coibentato in misura insufficiente, per raggiungere temperature fisiologicamente accettabili nei locali dell'ultimo piano, con la termoregolazione si realizzerebbe uno spreco di energia termica nei piani inferiori).

Ancora più efficaci risultano gli interventi sulle dispersioni termiche. Una maggiore coibentazione, in particolare del sottotetto e l'installazione di doppi vetri alle finestre sono provvedimenti applicabili anche a costruzioni già esistenti. La efficacia della prima azione, misurata in termini di risparmio percentuale di combustibile, sarà tanto maggiore quanto minore è il numero di piani dell'edificio: nelle diverse condizioni climatiche italiane, il risparmio oscilla fra il 20 e il 25 per cento per una villetta tipo, fra il 18 e il 20 per cento per un edificio a due piani con tre appartamenti medi per piano, e rimane ancora fra il 6 e il 10 per cento nel caso di uno stabile di 9 piani con tre appartamenti medio-grandi per piano. In nuove costruzioni il risparmio conseguibile con un buon isolamento termico, un rapporto ottimale vuoti/pieni, particolari accorgimenti nella progettazione (*si veda la figura a pagina 22*), un efficiente impianto di riscaldamento, potrebbe così arrivare mediamente al 45-50 per cento. Un'altra voce rilevante, classificata allo interno della voce «altri settori» nella tabella a pagina 21 su cui agire per una contrazione della domanda di energia, è data certamente dagli elettrodomestici. Ragionando in termini di rendimento $\eta$, questo risulta molto variabile (di un fattore due e anche oltre) da un apparecchio all'altro: il solo motore elettrico presenta sovente rendimenti inferiori al 50 per cento: se - oltre ad avvalersi di motori con elevato rendimento intrinseco - certi elettrodomestici (lavatrice, lavapiatti) fossero in grado di recuperare energia termica dall'acqua calda di scarico, potrebbero ulteriormente ridurre la domanda di energia.

Nel settore industriale occorre innanzitutto distinguere fra interventi sul processo produttivo e interventi sui prodotti finali di tali processi, in questo secondo caso al fine di ridurre l'uso di materiali ottenibili solo con procedimenti a elevata intensità di energia e/o l'obsolescenza del prodotto stesso o di sue parti. Un esame dettagliato di tutte le possibili proposte è impossibile, data la ampiezza e l'articolazione del campo di indagine. Globalmente parlando, una riduzione della domanda di energia del 10-15 per cento è ritenuta possibile con progetti intesi a realizzare maggiori rendimenti $\eta$ e con una più accurata manutenzione, in particolare di caldaie e di scambiatori di calore. Gli interventi più convenienti sono quelli nei settori industriali a maggiore intensità di energia, sia per migliorarne il rendimento dal punto di vista energetico, sia al fine di spostare per quanto possibile lo sviluppo produttivo verso i comparti a minore domanda energetica. La tabella a pagina 23, che si riferisce alla situazione italiana, illustra il divario esistente tra comparto e comparto, e non soltanto in termini energetici. Infatti non solo l'intensità energetica (definita come «consumi» energetici totali per unità di valore aggiunto) varia notevolmente - di quasi un fattore 250 fra il minimo e il massimo -, ma fra intensità energetica di un settore e investimento per addetto esiste una notevole correlazione, così che le attività a elevata domanda di energia risultano anche quelle a elevata intensità di capitale e, per converso, ad attività con elevata occupazione corrisponde una bassa intensità energetica. Non si tratta di un risultato occasionale. Diversi autori, partendo da campi di indagine non necessariamente omogenei ed elaborando ipotesi di sviluppo economico ed energetico fra loro dissimili per obiettivi o strumenti adottati, sono arrivati alla medesima conclusione, relativamente al rapporto che di norma esiste fra produzione a bassa intensità di energia ed elevata occupazione.

Ultimo nel tempo, e particolarmente significativo per le dimensioni dell'impegno, che generalizza l'analisi all'intero sistema, il piano di sviluppo energetico elaborato da un gruppo di scienziati danesi per il proprio paese e basato su una razionale gestione delle risorse di energia: rispetto al piano governativo, fra il 1976 e il 1995, senza mutare nella sostanza l'assetto economico e sociale, a un risparmio di fonti di energia non rinnovabili del 15,5 per cento corrisponderebbe un incremento dell'occupazione pari al 5,4 per cento. Questo fatto costituirebbe un incentivo anche sociale per la promozione di un sistema produttivo capace di contenere la propria domanda di energia.

Per quanto concerne viceversa il settore dell'industria dell'energia, un'azione basata sul miglioramento del rendimento $\eta$ non sembra in grado di produrre effetti molto significativi. Il rendimento di una centrale termoelettrica è oggi stabilizzato intorno a 0,40, mentre per le centrali nucleari oggi commerciali è ancora più basso (di poco superiore a 0,30), con la possibilità di salire a 0,40 se e quando diverranno commerciali le centrali basate su reattori autofertilizzanti veloci. La restante parte dell'energia in ingresso viene dissipata sotto forma di calore a bassa temperatura. Incrementi ulteriori (fino a $\eta=0,50$ e forse oltre) sarebbero possibili mediante cicli combinati gas-vapore o con la generazione magnetoidrodinamica. Le difficoltà pratiche (sia tecnologiche che economiche) da superare affinché queste soluzioni abbiano successo, sono però tali da rendere problematica, e comunque proiettata in un lontano futuro, la loro adozione su

| FORME NELLE QUALI È FORNITA L'ENERGIA | PERCENTUALE SUL TOTALE | | | RIPARTIZIONE DEL CALORE FORNITO SECONDO LE TEMPERATURE (IN PERCENTUALE DEL CONSUMO TOTALE DI ENERGIA) | | | |
|---|---|---|---|---|---|---|---|
| | USA | CANADA | REGNO UNITO | | USA | CANADA | REGNO UNITO |
| CALORE | 58 | 69 | 65 | CALORE FORNITO: | | | |
| LAVORO MECCANICO | 38 | 27 | 30 | — A TEMPERATURE SUPERIORI A 100 °C | 24 | 30 | 10 |
| — TRASPORTI | (34) | (24) | (27) | | | | |
| — FORZA MOTRICE | (4) | (3) | (3) | — A TEMPERATURE INFERIORI A 100 °C | 34 | 39 | 55 |
| ELETTRICITÀ | 4 | 4 | 5 | TOTALE CALORE | 58 | 69 | 65 |

**I dati di tabella, che si riferiscono all'anno 1973, mettono in evidenza come qualcosa fra i tre quinti e i due terzi dell'energia sia utilizzata sotto forma di calore, per di più in gran parte a temperature relativamente basse. Eppure, nell'attuale organizzazione della società, questo calore viene quasi sempre messo a disposizione bruciando combustibili pregiati, in grado di per sé di fornire calore a temperature molto più elevate. È questa la forma di «spreco» energetico a prima vista meno evidente, ma più rilevante e anche più difficile da rimuovere in quanto comporta una complessa, e non sempre incoraggiata riorganizzazione del sistema di produzione e di consumo delle diverse forme di energia.**

**In questo schema il valore estratto da una turbina a contropressione produce acqua calda a temperatura elevata mediante scambiatori di calore (*non illustrati in figura*). Nei periodi di punta della domanda termica entra in funzione anche un inceneritore, che fornisce l'ulteriore quantità di acqua calda richiesta (se non bastasse, si aggiungono di norma delle caldaie tradizionali). Per garantirsi dalle probabili fluttuazioni nella disponibilità del combustibile o nel suo prezzo, la caldaia principale, che produce il vapore ad alta pressione per la turbina, è predisposta per funzionare tanto a metano quanto a gasolio. Il primo viene comunque sempre utilizzato per gli usi domestici di cucina.**

scala significativa. Per quanto riguarda poi il trasporto dell'energia elettrica, da decenni si assiste allo sviluppo di linee a tensione crescente al solo fine di compensare - dal punto di vista delle perdite di energia - la parallela crescita della quantità di energia da trasportare (e delle relative linee di trasporto): pertanto con rendimenti complessivi sostanzialmente immutati. Miglioramenti importanti in assoluto possono infine aversi nella prestazione energetica delle raffinerie di petrolio, come pure nel trasporto a distanza del carbone e del gas naturale, ma non tali da modificare in misura significativa il contributo percentuale del settore delle industrie di energia, dove acquistano pertanto particolare evidenza i limiti di un approccio basato esclusivamente su incrementi del rendimento $\eta$.

Rimane il settore dei trasporti, dove un risparmio in fonti energetiche unitario e complessivo è realizzabile essenzialmente nel campo del trasporto su strada. Motori meno «spinti» (per esempio con un ridotto rapporto di compressione) possono operare con rendimenti superiori agli attuali. L'orientamento verso veicoli di potenza relativamente modesta porta nella medesima direzione. Inoltre miglioramenti nel profilo aerodinamico possono incrementare il rendimento del 10-15 per cento soprattutto alle elevate velocità. Anche i motori Diesel consentono rendimenti del 20-30 per cento superiori a quelli dei motori a scoppio e incrementi notevoli nel rendimento si hanno passando a motori a carica stratificata. Una percentuale non molto inferiore al 50 per cento dell'energia meccanica trasmessa dall'asse del motore viene mediamente dissipata durante le frenate: il recupero dell'energia cinetica può per esempio avvenire con un volano meccanico. Un altro criterio per aumentare il rendimento $\eta$ è quello di accoppiare al motore un accumulatore, che consenta un esercizio del motore a potenza costante al variare della potenza richiesta alle ruote: si tratta cioè di sviluppare un veicolo a motore ibrido, in cui l'energia sia immagazzinata nei periodi di bassa potenza richiesta sotto forma elettrica in un apposito accumulatore, il quale alimenterà un motore elettrico nei momenti di punta della domanda. L'utilizzazione di combustibili diversi da quelli tradizionali (per esempio miscele benzina-alcool metilico) sembra infine in grado di consentire la realizzazione di rendimenti più elevati degli attuali. Anche senza modificare la struttura del traffico è pertanto possibile ottenere risparmi notevoli nella domanda di energia.

Quando viceversa si voglia affrontare il problema del risparmio di fonti energetiche, oltre che migliorando $\eta$, anche e soprattutto agendo su $\varepsilon$, esso non può più in linea di principio essere risolto, considerando separatamente il singolo settore indicato nella tabella a pagina 21. Infatti la ricerca del rendimento massimo $\varepsilon$ presuppone per sua definizione l'organizzazione di diversi sistemi di trasformazione energetica, ciascuno avente il massimo rendimento $\eta$ compatibile con la funzione che deve svolgere, in modo tale che i sistemi così organizzati richiedano un contenuto di informazione il più possibile vicino a quello offerto dalla fonte energetica che si intende usare. È allora evidente che i singoli sistemi dovranno essere prescelti indipendentemente dalla loro appartenenza a questo o quel settore. Essi dovranno inoltre risultare il più possibile «integrati» dal punto di vista territoriale al fine di ridurre le perdite di energia dovute al suo trasporto. E, per lo stesso motivo, la generazione delle diverse forme di energia dovrà collocarsi nel baricentro del macrosistema costituito dai singoli sistemi utilizzatori.

Le suddette esigenze contrastano con le tendenze storiche nella organizzazione delle attività umane, in particolare con la specializzazione delle aree e con le opposte soluzioni di una centralizzazione spinta nella generazione di energia elettrica e di una decentralizzazione spinta nella generazione di energia termica. D'altra parte l'obiettivo di massimizzare $\varepsilon$ non contrasta affatto con l'attuale ripartizione degli usi finali dell'energia nei paesi industrializzati, come dimostrano gli esempi riportati nella tabella a pagina 24. Mentre i fabbisogni di energia elettrica non sostituibili con altre forme di energia sono modesti, l'energia termica - in particolare l'energia termica a bassa temperatura - è la forma di energia per cui sussiste la massima domanda finale. Orbene, si tratta proprio della forma di energia in larga misura dissipata senza alcuna utilità nel processo di produzione di energia elettrica. È quindi immaginabile una generazione di energia elettrica con sistemi parzialmente in scala ridotta rispetto agli attuali, da cui si ricavi inoltre energia termica in condizioni termodinamiche tali da consentire la sua applicazione in processi industriali e quindi - ulteriormente degradata dal punto di vista termodinamico - nel riscaldamento di edifici e in altri usi domestici. Si ha insomma quella che viene definita una «cascata di energia». Inoltre, una modesta quota dell'energia elettrica può alimentare pompe di calore portando così a temperature utili (adatte per esempio per il riscaldamento domestico) le risorse termiche disponibili nell'ambiente. Un concetto di questo genere, definito a «energia totale», oltre ad avvalersi dei vantaggi di singoli sistemi a elevato rendimento $\eta$ (che già ridurrebbero la domanda di calore illustrata nella tabella a pagina 24) e delle ridotte perdite di energia durante il trasporto, si integra molto bene con l'uso attivo e non solo passivo dell'energia solare (collettori solari come fonte integrativa di energia termica). Non solo, ma la pianificazione territoriale resa necessaria da simili sistemi può di per sé contribuire a migliorare il rendimento dell'insieme. Per esempio, uno studio recente riferito a una situazione concreta (in Alaska) ha messo in evidenza un 25 per cento circa di risparmio di combustibile adottando i sistemi a energia totale, mentre un risparmio pressoché identico può derivare dalla pianificazione urbanistica che consente un orientamento degli edifici, un loro distanziamento reciproco, ecc., tale da ottimizzare lo sfruttamento passivo della radiazione solare.

L'illustrazione a pagina 25 mostra uno schema pratico che approssima lo schema concettuale appena descritto, mentre gli schemi in questa pagina si riferiscono al caso specifico di «cogenerazione» di energia elettrica e termica in processi industriali.

Un altro strumento per la massimizzazione di $\varepsilon$ consiste nel recupero dei rifiuti solidi e liquidi di natura sia organica che inorganica. Per quelli incombustibili, in particolare di natura metallica, il loro riciclo richiede di norma una quantità di energia piuttosto ridotta rispetto alla produzione ex-novo: tipicamente, nel caso dei metalli non ferrosi l'energia richiesta è un quinto. Eppure, malgrado gli evidenti vantaggi energetici, il riciclo dei rottami dei principali prodotti metallici è assai modesto. Per esempio negli Stati Uniti si ricicla solo il 31 per cento dell'acciaio, il 39 per cento del piombo, zinco e rame, il 18 per cento dell'alluminio e il 20 per cento del nichel. Si tratta di valori non elevati, anche se, per riciclare convenientemente un materiale, occorre che per effetto del primo impiego

**La «cogenerazione», che consiste nella produzione combinata di energia elettrica ed energia termica per usi industriali, può venire schematicamente illustrata dalle figure in queste due pagine. Nella figura qui in alto, superiormente è indicato il processo nella forma tradizionale, mentre inferiormente viene illustrata la cogenerazione delle due forme di energia: in questo caso le caratteristiche termodinamiche del vapore richiesto (pressione: 3,5 kg/cm², temperatura: 150 °C) sono tali da consigliare di estrarlo dall'ultimo stadio di una turbina a contro pressione, a valle quindi della sua utilizzazione per generare l'energia elettrica. Il caso opposto si ha nelle figure in questa pagina, dove i gas ad alta temperatura prodotti nel processo industriale vengono trasferiti in una caldaia, nella quale si genera vapore ad alta pressione da inviare nella turbina collegata al generatore elettrico: in questo caso è l'energia elettrica a essere prodotta a valle.**

**La digestione anaerobica è il metodo di gran lunga più diffuso per la produzione di metano da liquami. I microrganismi anaerobici metabolizzano, mediante processi di riduzione, i composti organici. Come nei processi anaerobici, gli organismi usano azoto, fosforo e altre sostanze nutritive per sviluppare il protoplasma cellulare, e riducono l'azoto organico ad acidi organici e ammoniaca. La frazione di carbonio inutilizzata è liberata sotto forma di metano e, in piccola parte, di anidride carbonica. Il gas così prodotto può venire variamente sfruttato.**

esso non sia stato disperso nell'ambiente (come accade per il piombo aggiunto alla benzina) o non abbia subito trattamenti tali da renderlo «irreversibile».

Per i rifiuti organici, la scelta di una loro combustione è da giudicarsi come un'ipotesi riduttiva dal punto di vista della massimizzazione di $\varepsilon$. Le molecole delle sostanze organiche presentano infatti una organizzazione a elevato contenuto d'informazione, che va pertanto conservato, per quanto possibile, durante la trasformazione dei rifiuti. Di qui l'opportunità di riciclare carta e stracci e di utilizzare altri rifiuti (alimentari, umani, animali) per la generazione di combustibili pregiati (metano) mediante fermentazione anaerobica e la produzione in alternativa o in cascata alla precedente di sostanze adatte a rinnovare l'humus dei terreni agricoli, riducendo così il fabbisogno dei fertilizzanti tradizionali. Per i rifiuti residui, la loro combustione diretta (o dei prodotti di un processo di pirolisi) dovrebbe servire alla produzione combinata di energia elettrica e termica, come integrazione di quella generata con combustibili tradizionali, entrando così a far parte attiva di un sistema a energia totale. D'altronde la convenienza energetica (ed economica) a raccogliere, riciclare o bruciare i rifiuti è strettamente legata alla minimizzazione della distanza a cui tali rifiuti vanno trasportati: e quindi la generalizzazione di queste pratiche comporta proprio lo sviluppo di un'organizzazione decentrata e integrata della produzione e del consumo, già richiesta dai sistemi a energia totale. Questi concetti, sono tutti applicabili anche nel settore agricolo, come è illustrato nella figura qui in alto.

La disponibilità di liquami organici provenienti dagli allevamenti di animali, e la possibilità di ricavarne metano, si sposano con gli indubbi vantaggi di sistemi energetici autonomi in zone decentrate. Il metano così prodotto può essere allora utilizzato per alimentare un impianto su piccola scala per la generazione congiunta di energia elettrica e termica. Un impianto del genere è per esempio il TOTEM, costituito sostanzialmente da un motore di una FIAT 127, alimentato a metano e accoppiato con un generatore elettrico da 15 chilowatt. Il calore residuo viene recuperato attraverso una serie di scambiatori di calore, così da arrivare a un rendimento $\eta$ massimo superiore a 0,9. Una soluzione siffatta è in fase di sviluppo presso Reggio Emilia, in un'azienda agricola con un significativo allevamento di suini.

Sempre con riferimento alla tabella a pagina 24, una larga quota di energia finale è richiesta sotto forma di lavoro meccanico per il trasporto. Oltre alle riduzioni in tale domanda di energia derivabili da incrementi nel rendimento $\eta$ (valutabili fra il 30 e il 50 per cento rispetto a una situazione basata sui sistemi attuali), va sottolineato che una struttura integrata e decentrata contribuirebbe di per sé a minimizzare l'onere dei trasporti. Se a questo si aggiungesse, com'è auspicabile, un riorientamento verso mezzi collettivi di trasporto, la domanda di energia - sotto forma di combustibili ad alto costo quali i derivati leggeri del petrolio - potrebbe risultare ulteriormente contratta.

Il risultato finale di una ipotetica scelta strategica intesa a massimizzare il rendimento $\varepsilon$ complessivo di un sistema energetico, potrebbe essere tale da sconvolgere i più accreditati luoghi comuni: per esempio che lo sviluppo economico di un paese richiede necessariamente la disponibilità di una maggiore quantità di energia. Sembra un'affermazione teoricamente inoppugnabile, anche se i dati della tabella in alto nella pagina a fronte mostrano il largo margine di manovra esistente. Al di là di questi dati empirici, che dovrebbero comunque suggerire una certa prudenza, resta il fatto che l'affermazione dell'esistenza di un rapporto di causa ed effetto fra crescita dell'economia e crescita della domanda di energia discende anch'essa da una concezione delle trasformazioni energetiche basate sul primo principio della termodinamica. Se alla maggior domanda di energia finale richiesta da una nuova produzione o da un nuovo servizio si risponde solo cercando di migliorare i rendimenti $\eta$ dei sistemi energetici a ciò preposti, di norma ci si limiterà a ridurre l'incremento di fonte energetica primaria rispetto al caso di mancato intervento sui rendimenti $\eta$. Se, viceversa, si agisse per aumentare in modo continuo il rendimento $\varepsilon$ dell'intera organizzazione sociale, sarebbero attuabili drastiche riduzioni nella domanda di fonti energetiche primarie non rinnovabili a parità di energia utile finale. È quanto avviene per esempio nel caso della figura in basso a destra. Tutto questo, evidentemente, ha un costo. Lo sviluppo di sistemi a energia totale comporta un incremento nei beni strumentali (generatori di energia di vario tipo, scambiatori di calore, tubazioni, sistemi solari, ecc.) rispetto al caso di un'organizzazione energetica meno decentrata. Anche se la maggior domanda di tali beni potrà portare a riduzioni di costo per effetto della loro produzione in serie, appare difficile immaginare una diminuzione degli investimenti complessivi richiesti; anzi, è ragionevole attendersi un loro incremento. Questo elemento va però valutato correttamente. Innanzitutto si può già oggi osservare in diversi casi che i maggiori costi di investimento sono compensati (o più che compensati) dai minori costi di approvvigionamento di fonti energetiche non rinnovabili. Più in generale, poiché una strategia di trasformazione radicale dell'assetto energetico avrà effetto sul medio-lungo periodo, va tenuto conto del fatto che le previsioni più attendibili concordano oggi nel ritenere inevitabili ulteriori aumenti in termini reali del costo delle risorse energetiche non rinnovabili: di conseguenza sul medio-lungo periodo maggiori costi d'investimento potrebbero risultare comunque convenienti rispetto a ulteriori incrementi nel costo dei combustibili. In terzo luogo già oggi, per lo meno per i paesi importatori di energia (cioè per tutti i paesi europei), la riduzione nella domanda di fonti primarie non rinnovabili mediante sistemi a energia totale comporta una diminuzione dell'esborso valutario (per acquistare energia dall'estero) a favore di un incremento del valore aggiunto prodotto all'interno, e quindi anche dell'occupazione. Poiché gli stati moderni sopportano costi notevoli per intervenire a favore dei disoccupati, la conseguente riduzione di tali oneri andrebbe pertanto conteggiata nell'analisi costi-benefici; per non parlare degli evidenti vantaggi anche sociali di un incremento dell'occupazione e di una riduzione nel deficit valutario.

Si ritrovano insomma su scala globale e per deduzione razionale le stesse indicazioni di carattere sociale emergenti dai dati della tabella a pagina 23, a conferma delle implicazioni ampie, non solo sotto il profilo fisico, di una corretta politica di conservazione delle fonti di energia. La quale, tuttavia, va vista nel contesto di un più ampio disegno di sviluppo economico e sociale. Solo in quest'ambito sarà possibile determinare le dimensioni e la scala dei tempi dei tipi di interventi qui delineati.

| PAESE | 1950-1960 | 1960-1970 |
|---|---|---|
| RDT | 0,60 | 0,83 |
| SPAGNA | 0,99 | 0,97 |
| POLONIA | 0,85 | 0,83 |
| URSS | 0,79 | 0,88 |
| RFT | 0,74 | 0,96 |
| REGNO UNITO | 0,90 | 0,88 |
| CECOSLOVACCHIA | 0,86 | 0,92 |
| BELGIO | 0,85 | 0,98 |
| IUGOSLAVIA | 0,84 | 0,99 |
| FRANCIA | 0,88 | 1,01 |
| AUSTRIA | 0,90 | 1,04 |
| ISLANDA | 1,21 | 0,85 |
| UNGHERIA | 1,25 | 0,82 |
| ROMANIA | 1,14 | 1,00 |
| OLANDA | 1,06 | 1,21 |
| PORTOGALLO | 1,10 | 1,09 |
| NORVEGIA | 1,14 | 1,20 |
| DANIMARCA | 1,08 | 1,39 |
| SVEZIA | 1,36 | 1,25 |
| SVIZZERA | 1,33 | 1,35 |
| GRECIA | 1,84 | 1,12 |
| FINLANDIA | 1,45 | 1,63 |
| BULGARIA | 1,73 | 1,49 |
| ITALIA | 2,00 | 1,45 |

**Il coefficiente di elasticità, definito come l'incremento percentuale di domanda di energia richiesto per aumentare dell'1 per cento il prodotto nazionale lordo, varia notevolmente per i differenti paesi europei e in due diversi decenni recenti. I dati della tabella smentiscono in particolare la teoria secondo cui il coefficiente di elasticità dovrebbe decrescere col tempo nei paesi già sviluppati: nella Repubblica Federale Tedesca e in Francia si ha per esempio un incremento del coefficiente di elasticità nel passaggio dal decennio 1950-1960 al decennio 1960-1970.**

**Se si ipotizzano tre modelli di crescita nella domanda di energia primaria per la CEE sino al 2000, basati rispettivamente su un «modello inerziale» (proiezione nel futuro del modo tradizionale di trasformare e utilizzare energia), su una politica d'intervento sui soli rendimenti $\eta$, su una politica tesa a un graduale e continuo incremento del rendimento $\varepsilon$, pur ammettendo in tutti e tre i casi lo stesso incremento annuo nella popolazione (0,8 per cento), nel numero di autovetture (3,6 per cento), nel prodotto nazionale lordo (4,0 per cento), la figura mette in evidenza la possibilità di uno sviluppo economico e sociale basato su una riduzione nella domanda di energia primaria. Analoghi risultati hanno conseguito recenti studi compiuti in Germania e negli USA.**

# La sostituzione chirurgica dell'articolazione del ginocchio

*Grazie a recenti miglioramenti nelle tecniche di progettazione, nei materiali e nei metodi chirurgici, il ginocchio artificiale assicura vantaggi sostanziali ai pazienti colpiti da gravi forme di artrite*

di David A. Sonstegard, Larry S. Matthews e Herbert Kaufer

Nel corso di attività normali come il camminare, il correre, l'inginocchiarsi, il salire le scale e il sedersi o l'alzarsi da una sedia, il carico che viene imposto all'articolazione del ginocchio dell'uomo può superare di cinque volte il peso del corpo. Oggi, molte persone pretendono dal ginocchio prestazioni supplementari (spesso eccessive), praticando sport quali il football, il calcio americano, il tennis, la corsa di fondo. Non desta quindi meraviglia il fatto che molte persone arrivino agli anni della vecchiaia con un ginocchio, o addirittura con entrambi, così mal ridotti da essere costretti a un'andatura zoppicante.

È oggi possibile sostituire il ginocchio così rovinato con un dispositivo meccanico che imita i raffinati e complessi movimenti di cui un ginocchio sano è capace. Fino a ora la protesi serve soltanto a persone gravemente menomate e che, verosimilmente, non pretendono dal ginocchio grandi prestazioni. Il ginocchio artificiale non costituisce tuttavia una risposta ai problemi di coloro la cui attività è ancora di tipo giovanile, soprattutto perché i carichi cui essi sottopongono il ginocchio non possono essere sopportati dalla protesi artificiale, sulla base delle tecniche attualmente disponibili per la sua fissazione alle ossa. In questo articolo descriveremo diversi tipi di ginocchio artificiale, ponendo l'accento soprattutto sul ginocchio «sferocentrico», messo a punto dal nostro gruppo presso l'Università del Michigan.

Il lavoro sul ginocchio artificiale è il risultato dei progressi ottenuti, in questi ultimi quindici anni, a proposito dell'articolazione artificiale dell'anca. Gran parte del merito relativo all'introduzione dei materiali e delle tecniche di progetto che dovevano servire per la realizzazione dell'articolazione dell'anca, e alla tecnica chirurgica in caso di sua applicazione, va al chirurgo inglese John Charnley del Centre for Hip Surgery, presso il Wrightington Hospital, nel Lancashire.

L'articolazione naturale dell'anca è un meccanismo del tipo del giunto a sfera che permette il movimento in tutti i piani passanti per il proprio centro di rotazione. (L'anca artificiale è progettata in modo analogo e permette un'analoga gamma di movimenti.) In più, l'articolazione dell'anca è circondata e protetta da una grande massa di tessuto muscolare e connettivo. Il ginocchio, al contrario, possiede una gamma di movimenti delimitata in modo più preciso e risulta scarsamente protetto dal tessuto circostante. È altamente suscettibile di azioni traumatiche causate sia da colpi, sia da arresti o da torsioni improvvisi. Per questi motivi, nel caso del ginocchio artificiale, non è stato facile riportare lo stesso successo ottenuto con l'anca artificiale. Anche l'anca artificiale, tuttavia, non è adatta a sostituire una articolazione gravemente danneggiata tranne che in casi molto particolari. Una frattura o un evento patologico altrettanto grave impediscono un appropriato allineamento così come un appropriato supporto scheletrico della protesi. Coloro, quindi, cui è stata impiantata una articolazione artificiale dell'anca o del ginocchio erano, in genere, pazienti invalidi affetti da artrite cronica.

Artrite significa letteralmente «infiammazione articolare». Il termine viene usato per indicare un centinaio circa di malattie, i cui sintomi comuni sono dolore persistente o ricorrente, rigidità, gonfiore e sensibilità eccessiva in una o più articolazioni. La Arthritis Foundation ha calcolato in 363 milioni il numero di individui affetti da artrite, vale a dire il dieci per cento della popolazione mondiale. Di questi, più di 50 milioni vivono negli Stati Uniti, tra i quali circa 20 milioni richiedono cure mediche e 3 milioni e mezzo sono gravemente invalidati.

La causa dell'artrite non è conosciuta e a tutt'oggi non esiste una terapia efficace per la maggior parte delle manifestazioni della malattia. Le forme principali sono la osteoartrite e l'artrite reumatoide. L'osteoartrite è un tipo di artrite degenerativa da usura, che progredisce in rapporto all'uso dell'articolazione e all'età del soggetto. L'artrite reumatoide, una malattia infiammatoria dei tessuti molli, è la forma più distruttiva e invalidante della malattia. Il trattamento di entrambe queste forme comprende la chemioterapia, la fisioterapia, il riposo, l'uso di fasciature e l'applicazione di calore. Nonostante questi tentativi, sia l'osteoartrite sia l'artrite reumatoide possono progredire fino a causare dolore intenso, instabilità dell'articolazione, deformazioni, limitazioni dell'attività e invalidità. È proprio questo il caso in cui si rende necessaria la terapia chirurgica. La maggior parte delle sostituzioni dell'articolazione dell'anca e del ginocchio vengono praticate in situazioni del genere.

È possibile comprendere a fondo quello che comporta la sostituzione dell'articolazione del ginocchio, considerando l'anatomia di un ginocchio normale (*si veda l'illustrazione a pagina 33*). La articolazione include due snodi articolari, ovvero parti reciprocamente movibili, che trasmettono la forza applicata e il movimento che ne deriva. Uno è costituito dalla superficie articolare rotuleo-femorale tra la rotula e il femore. L'altro è costituito dall'articolazione tibio-femorale tra la tibia (la maggiore delle due ossa lunghe situate tra il ginocchio e la caviglia) e il femore.

L'energia necessaria al funzionamento dell'articolazione del ginocchio è fornita da tre gruppi fondamentali di muscoli, sotto il controllo del sistema nervoso. Il gruppo del semitendineo, del semimembranoso e del bicipite femorale e quello



del gastrocnemio, contraendosi, permettono alla gamba di piegarsi articolandosi al ginocchio. Il gruppo del quadricipite, sfruttando l'appoggio meccanico della rotula, permette alla gamba di distendersi. Questi due movimenti sono conosciuti rispettivamente con il termine di flessione e di estensione. I legamenti del ginocchio, costituiti soprattutto da fibre parallele di collagene, sono flessibili, permettendo così una libertà di movimento considerevole, anche se notevolmente controllata. Lungo il loro decorso, i legamenti formano connessioni forti e relativamente inestensibili tra osso e osso, fornendo quindi all'articolazione stabilità e un vincolo alle possibilità di movimento. Il legamento patellare collega la rotula e la tibia, permettendo un movimento di scorrimento durante la flessione e la estensione a livello dell'articolazione rotuleo-femorale. Due paia di legamenti, i collaterali e i crociati, stabilizzano l'articolazione tibio-femorale.

L'articolazione tibio-femorale comprende due strutture di cartilagine fibrosa, a forma di mezzaluna: il menisco mediale (all'interno del ginocchio) e il menisco laterale (all'esterno). Le superfici superiori dei menischi sono concave, provvedendo un approfondimento delle superfici articolari della estremità superiore tibiale per i condili femorali, cioè per le due strutture terminali dell'estremità inferiore del femore, a forma di grosso tubercolo. Mezzi di unione con la superficie articolare della tibia, relativamente piatta, permettono il movimento dei menischi rispetto alla tibia. (Il comune trauma da attività sportiva, noto come «lacerazione del menisco» coinvolge uno o entrambi i menischi.)

I condili femorali mediale e laterale sono superfici complesse che presentano un raggio di curvatura costantemente variabile. Queste superfici articolari, assieme ai legamenti e sfruttando la forma geometrica dei menischi e della tibia, permettono al ginocchio una gamma di movimenti flessibile e ben controllata.

Le caratteristiche del movimento sono degne di nota. Ad esempio, nella flessione e nella estensione il ginocchio non agisce come un semplice ginglimo, ma piuttosto mediante una successione infinita di centri di rotazione singolarmente collocati, dove ciascun centro agisce secondo una particolare orientazione relativa del femore e della tibia. L'articolazione tibio-femorale permette quindi movimenti coordinati in modo complesso di tipo oscillatorio, scivolatorio e rotatorio. Alla massima estensione, una rotazione della tibia verso l'esterno induce un meccanismo di arresto che permette di assumere comodamente la stazione eretta.

**In queste radiografie è possibile vedere un ginocchio artificiale impiantato nella gamba sinistra di un paziente affetto da una grave forma di artrite. La protesi è del tipo «sferocentrico» realizzata dagli autori presso l'Università del Michigan, ove sono state impiantate oltre 120 protesi di questo tipo. Tra le caratteristiche distintive della protesi sferocentrica vi è il fatto che essa si avvicina al movimento triassiale di cui è capace il ginocchio normale e che permette una decelerazione graduale del movimento e non un arresto improvviso.**

Le superfici articolari della rotula, della tibia e del femore sono costituite da uno strato relativamente sottile di cartilagine, un tessuto di interposizione. La cartilagine è lubrificata dal liquido sinoviale, contenuto nello spazio articolare all'interno di una capsula. Sebbene vi sia ancora molto da imparare sulla lubrificazione naturale del ginocchio, si sa che la combinazione di liquido sinoviale e di tessuto cartilagineo dà luogo a un meccanismo levigato, durevole e a basso attrito, che la migliore articolazione artificiale non può eguagliare.

L'uso di protesi metalliche per la riparazione o per il sostegno dello scheletro umano è forse tanto antico quanto la chirurgia. Tuttavia l'organismo richiede al metallo prestazioni ardue. Fino al nostro secolo le esigenze di biocompatibilità, sterilità, forza e resistenza alla corrosione non erano adeguatamente soddisfatte. Solo pochi metalli si sono dimostrati adatti. Negli Stati Uniti la maggior parte delle articolazioni artificiali sono realizzate in acciaio inossidabile o in una lega di cobalto, cromo e molibdeno. Anche il titanio viene impiegato sempre più frequentemente.

Dall'inizio degli anni cinquanta, il lavoro di ricerca sui materiali adatti è progredito fino ad arrivare alla prima sostituzione del ginocchio, effettuata con notevole successo da Börje Walldius dell'Istituto Karolinska di Stoccolma. La sua protesi era progettata secondo uno schema a ginglimo fisso e fu all'inizio realizzata in materiale acrilico e successivamente in una lega di cobalto, cromo e molibdeno. La protesi viene fissata all'osso per mezzo di una accurata sagomatura e di adattamento dei componenti alle estremità della tibia e del femore e, per mezzo di sporgenze assili, che si prolungano nel cavo midollare diafisario di queste ossa lunghe. Il ginocchio di Walldius ha permesso la riduzione del dolore e il ripristino della funzionalità in numerosi pazienti artritici. Nei casi in cui si sono verificati degli insuccessi, le ragioni solitamente risiedevano nel distacco della protesi dall'osso o nella infiammazione dei tessuti a causa dei depositi derivati dall'usura per contatto delle superfici articolari (metallo contro metallo). Il progetto di base a ginglimo fisso, con alcune modifiche che riflettono il progresso tecnologico dei materiali impiegati nella realizzazione delle superfici articolari e dei mezzi di fissazione della protesi, è ancora preferito da diversi chirurghi per operare la sostituzione di un ginocchio gravemente invalidato.

Due importanti miglioramenti tecnologici sono stati ottenuti da Charnley nel suo lavoro sulla articolazione artificiale dell'anca. Il primo consiste nella realizzazione delle superfici articolari, da lui effettuata utilizzando metallo levigato insieme a polietilene ad alto peso molecolare. Questa combinazione fornisce una superficie articolare con un basso coefficiente di attrito, un basso tasso di usura e scarsi depositi da usura (depositi che, tra l'altro, sono inerti dal punto di vista biologico).

Il secondo miglioramento apportato da Charnley consiste nell'uso di una sostanza cementante, il polimetilmetacrilato, per la fissazione dei componenti della protesi all'osso. Il polimetilmetacrilato non è una colla né un adesivo ad azione chimica, ma un vero e proprio cemento che opera meccanicamente. Viene applicato allo stato di pasta, in modo da poter adattarsi plasticamente attorno ai rilievi e nelle impronte dell'osso; indurendosi esso forma un mezzo di fissazione di tipo meccanico.

Questi progressi sono alla base del successo dell'anca totale di Charnley. I due componenti che la costituiscono sono una cavità acetabulare sferica realizzata in polietilene e una testa femorale realizzata in metallo. Ciascun componente è fissato allo scheletro con polimetilmetacrilato.

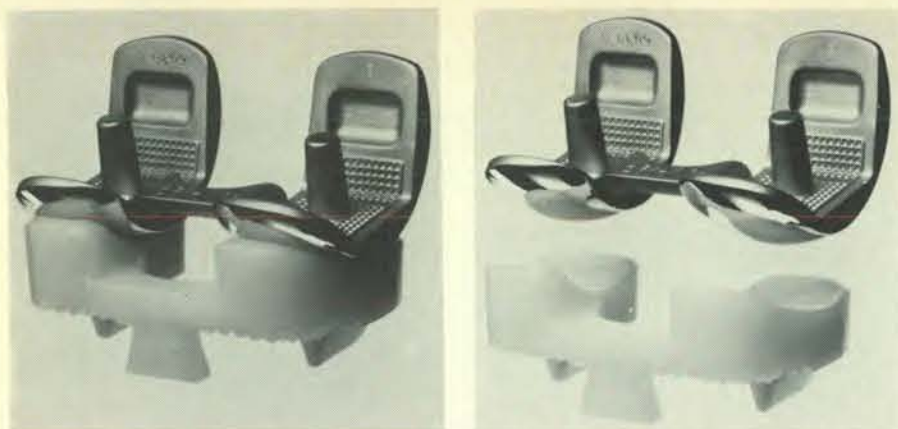
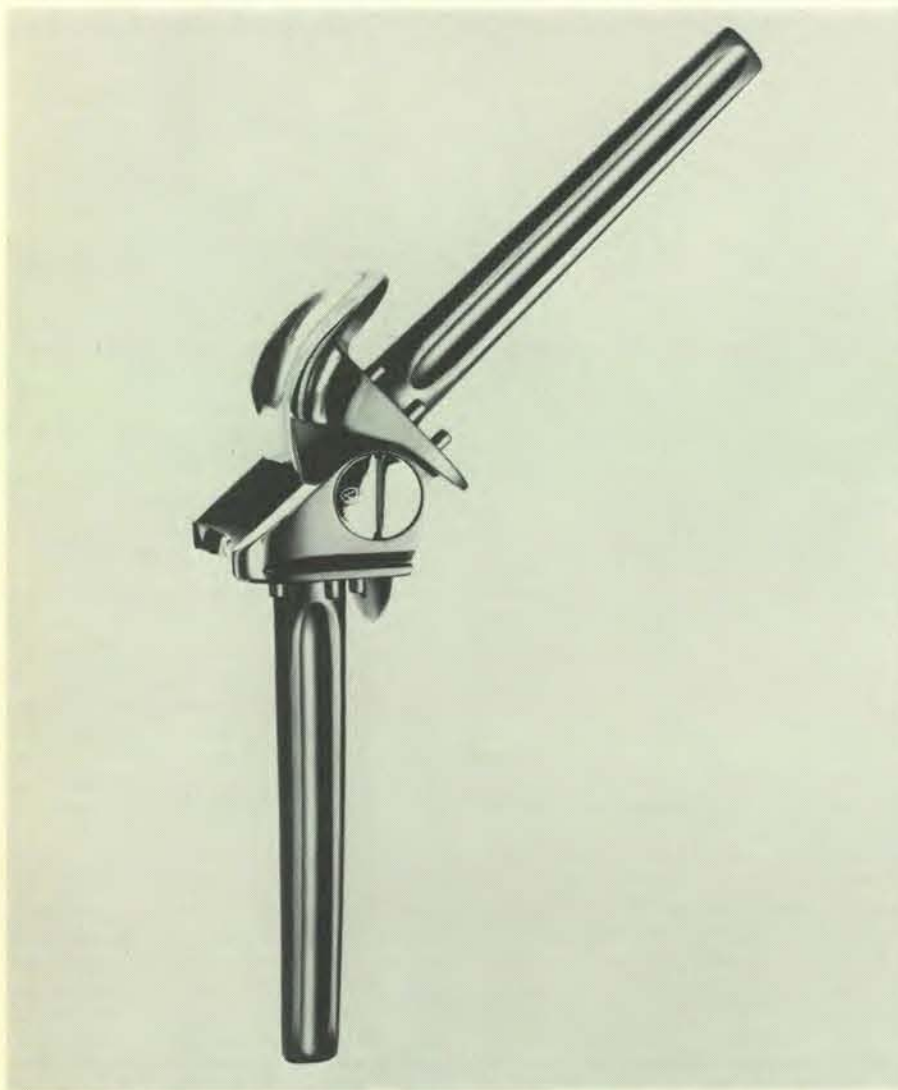
**Sono qui indicate le principali strutture scheletriche coinvolte nel funzionamento dell'arto inferiore dell'uomo. Nell'arto sinistro sono mostrate delle protesi. I progressi nel caso del ginocchio artificiale sono in parte dovuti al successo ottenuto con l'articolazione artificiale dell'anca.**

**Rappresentazione frontale e laterale del ginocchio sinistro normale. Nella prospettiva normale il ginocchio è un poco flesso, mostrando i recessi interni e la sistemazione dei legamenti. In quella laterale si vedono i principali gruppi muscolari che presiedono al funzionamento dell'articolazione. La flessione della gamba è realizzata dal gruppo dei muscoli semitendineo, semimembranoso e bicipite femorale e dal gruppo del gastrocnemio; l'estensione è realizzata dal gruppo del quadricipite. Nella prospettiva laterale parte del femore è resecata.**

Il primo progetto di ginocchio artificiale che sfruttò questi progressi fu proposto nel 1971 da Frank H. Gunston, ortopedico presso il Winnipeg General Hospital e l'Università di Manitoba. Il ginocchio di Gunston è una protesi «a binari e guide»: i «binari», realizzati in metallo sono fissati al femore e le «guide», realizzate in polietilene, sono cementate alla tibia. Le guide sono diritte dalla parte anteriore fino alla posteriore e leggermente concave da un lato all'altro; il movimento dei binari nelle guide è relativamente poco vincolato.

Negli ultimi anni è stato fatto molto lavoro per progettare un ginocchio artificiale ideale. I problemi in discussione comprendono la quantità di movimento relativo che deve esistere tra il componente femorale e quello tibiale, i materiali da usare per la realizzazione delle superfici articolari, l'integrità strutturale dei componenti realizzati in materiale plastico, il rimodellamento della superficie dell'articolazione rotuleo-femorale, la quantità di osso da rimuovere e i criteri di scelta dei pazienti. Si calcola che complessivamente nel mondo siano disponibili più di 80 progetti di ginocchio artificiale. Molti di essi si differenziano solo di poco, sia nel disegno che nella funzione, per cui una simile proliferazione può indurre in sospetto. Tuttavia, la situazione riflette gli sforzi intensi compiuti nella progettazione di protesi sostitutive del ginocchio.

Noi punteremo l'attenzione su due tipi generali di protesi sostitutive del ginocchio. La prima, che rappresenta il gruppo più vasto di protesi, si basa su un meccanismo non articolato o non vincolato. In questo caso l'accento è posto sulla sostituzione delle superfici articolari del ginocchio, usurate o malate. Tali protesi non prevedono un collegamento meccanico, un ginglimo o un analogo dispositivo articolare tra il componente femorale e quello tibiale. Al contrario la stabilità dell'articolazione è fornita dagli stessi legamenti e muscoli del paziente, sfruttando, in alcuni casi, le intrinseche caratteristiche geometriche della protesi. Il ginocchio di Gunston è un esempio in questo senso.

Il secondo tipo è la protesi articolata. Essa fornisce un collegamento meccanico vincolato tra il componente femorale e quello tibiale. Ne è un esempio classico il ginocchio a ginglimo fisso realizzato da

**Il ginocchio «geometrico» rappresenta una classe di ginocchi artificiali non articolati, vale a dire sprovvisti di un collegamento meccanico tra i componenti femorale e tibiale. La protesi è mostrata montata (*a sinistra*) e smontata (*a destra*). La stabilità di una protesi di questo tipo dipende dai legamenti e dai muscoli del paziente. In realtà, un ginocchio artificiale non articolato equivale a un rimodellamento delle superfici articolari del ginocchio del paziente. Il componente femorale del ginocchio geometrico possiede due superfici articolari realizzate in metallo; le corrispondenti zone concave del componente tibiale sono realizzate in polietilene.**

**La protesi di Walldius rappresenta uno dei primi ginocchi artificiali articolati ed è stata introdotta all'inizio degli anni cinquanta da Borje Walldius dell'Istituto Karolinska di Stoccolma. Il ginocchio di Walldius, progettato secondo lo schema di un ginglimo fisso, permette una gamma di movimenti, che è solo leggermente inferiore a quella di un ginocchio normale. Nella sua versione attuale il ginocchio viene realizzato in una lega di cobalto, cromo e molibdeno.**

Walldius del quale abbiamo già parlato.

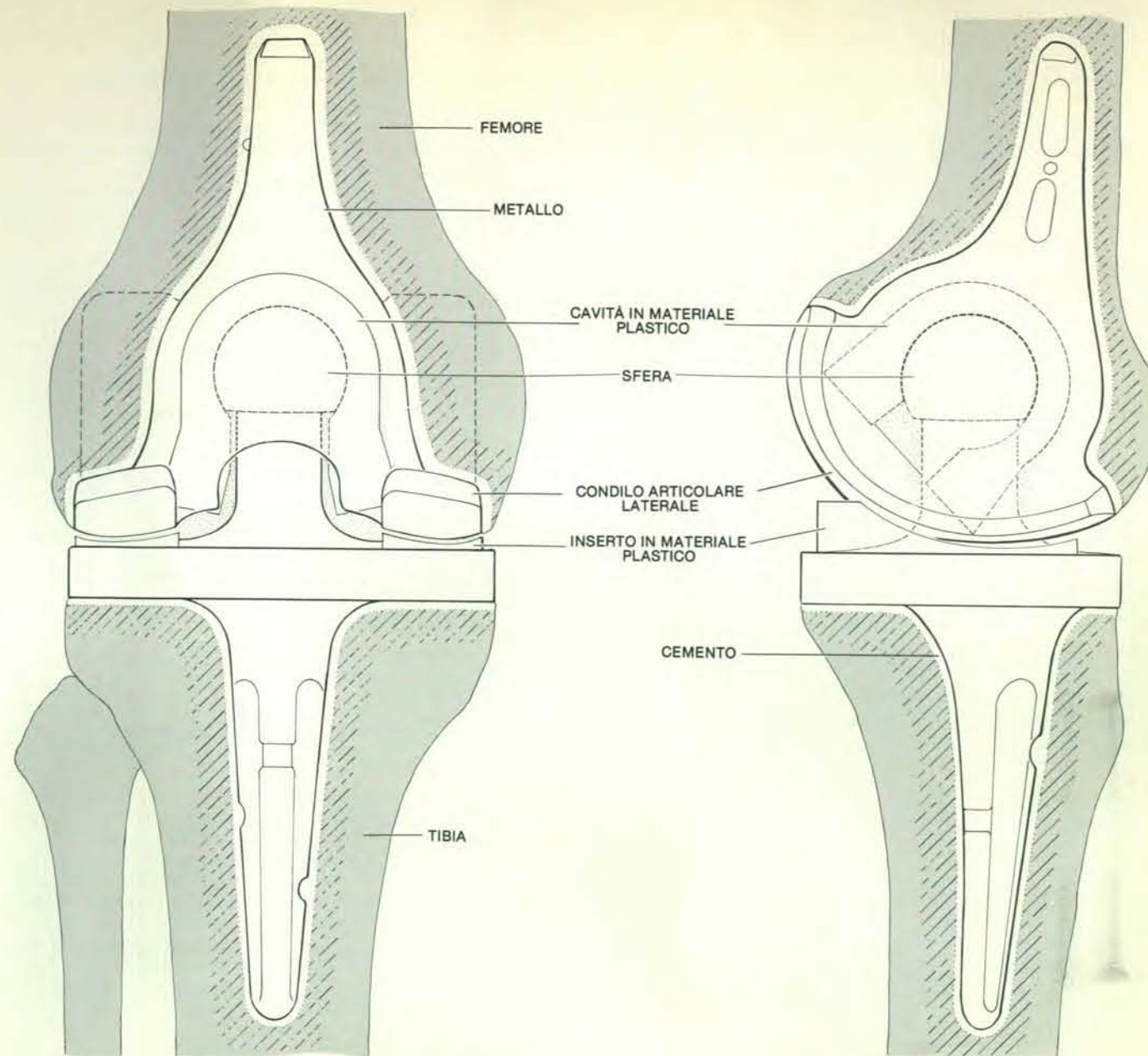
Una delle protesi non articolate, più comunemente adottate negli Stati Uniti, è il ginocchio «geometrico», che è il risultato del lavoro di ricercatori ortopedici di diverse istituzioni. Dal punto di vista concettuale esso è simile al ginocchio di Gunston, ma possiede due componenti anziché quattro. Il chirurgo si trova quindi a dover risolvere un minor numero di problemi di inserimento e di allineamento. Le superfici articolari sono più congrue, permettendo quindi zone portanti di carico più ampie e un certo vincolo dei movimenti.

Il componente femorale consiste di due superfici articolari realizzate in metallo e fissate ai condili (il metallo è una lega di cobalto, cromo e molibdeno) e aventi la forma di sfere tronche. Il componente tibiale, realizzato in polietilene, presenta concavità sferiche per l'alloggio dei condili, di diametro appena superiore. La protesi permette quindi una flessione di 130 gradi (il ginocchio normale è capace di una flessione di circa 135 gradi, quindi di solo di poco superiore) e piccoli movimenti rotatori laterali e assiali.

Entrambi i componenti possiedono un ponte che collega le due metà simmetriche attraverso il piano frontale. Ciascun componente quindi ha, visto dall'alto, la forma di una *U*. Questa caratteristica permette al chirurgo di conservare i legamenti crociati e collaterali. (In caso di sostituzione totale, l'anatomia del ginocchio rende possibile l'installazione della protesi senza che si debba ricorrere alla rimozione dei muscoli.)

Per impiantare un ginocchio geometrico, il chirurgo rimuove dapprima un sottile strato delle superfici articolari del ginocchio malato. Quindi le estremità delle ossa vengono sagomate in modo da ottenere il giusto allineamento e la corretta sistemazione dei componenti della protesi, che vengono fissati per mezzo di polimetilmetacrilato. I componenti della protesi possiedono sporgenze, solchi e incastri a coda di rondine, progettati per rendere più salda l'azione del cemento.

Il paziente con un ginocchio geometrico possiede quindi l'equivalente di un rifacimento delle superfici articolari del suo ginocchio artritico. In circa 10 anni di applicazioni, il ginocchio geometrico ha dimostrato di poter fornire sollievo al dolore e di restituire al paziente un certo grado di attività, certamente maggiore di quanta non ne potesse esplicare prima dell'intervento. Le protesi si sono dimostrate insufficienti solo in pochi casi, a causa soprattutto del distacco del componente tibiale o di una sua deformazione. In casi del genere è spesso possibile impiantare un'altra protesi, come il ginocchio «sferocentrico».

Il ginocchio sferocentrico è il risultato del lavoro che il nostro gruppo iniziò nel 1971, con lo scopo di progettare un ginocchio artificiale articolato. Le protesi non articolate come il ginocchio geometrico svolgevano bene la loro funzione in quei pazienti i cui problemi riguardavano soprattutto le superfici articolari che conservavano legamenti in buono stato. Nel caso di un ginocchio gravemente invalidato dall'artrite, con gravi deformazioni e instabilità, la protesi a ginglimo fisso, del tipo Walldius, era la sola a essere comunemente usata. Considerando il continuo progresso delle protesi sostitutive, pensammo che un'alternativa fosse opportuna e possibile.

Ci proponemmo alcune mete e alcuni criteri di progetto, tra i quali quello di ottenere una interconnessione articolata, di tipo meccanico, tra i componenti femorale e tibiale, in modo da fornire una stabilità intrinseca indipendentemente dai legamenti; di dotare la protesi di un movimento su tre assi, caratteristico del ginocchio normale; di delimitare i movimenti possibili mediante una decelerazione graduale piuttosto che con un arresto improvviso; di contrapporre metallo a polietilene a livello di tutte le superfici articolari; di rinforzare tutte le parti in polietilene con metallo allo scopo di minimizzare la tendenza del materiale plastico alla deformazione; di utilizzare il polimetilmetacrilato per la fissazione delle protesi; di assicurare la sostituzione di qualunque superficie articolare in polietilene, in caso di usura, e di ottenere una protesi che il chirurgo potesse impiantare senza soverchie difficoltà.

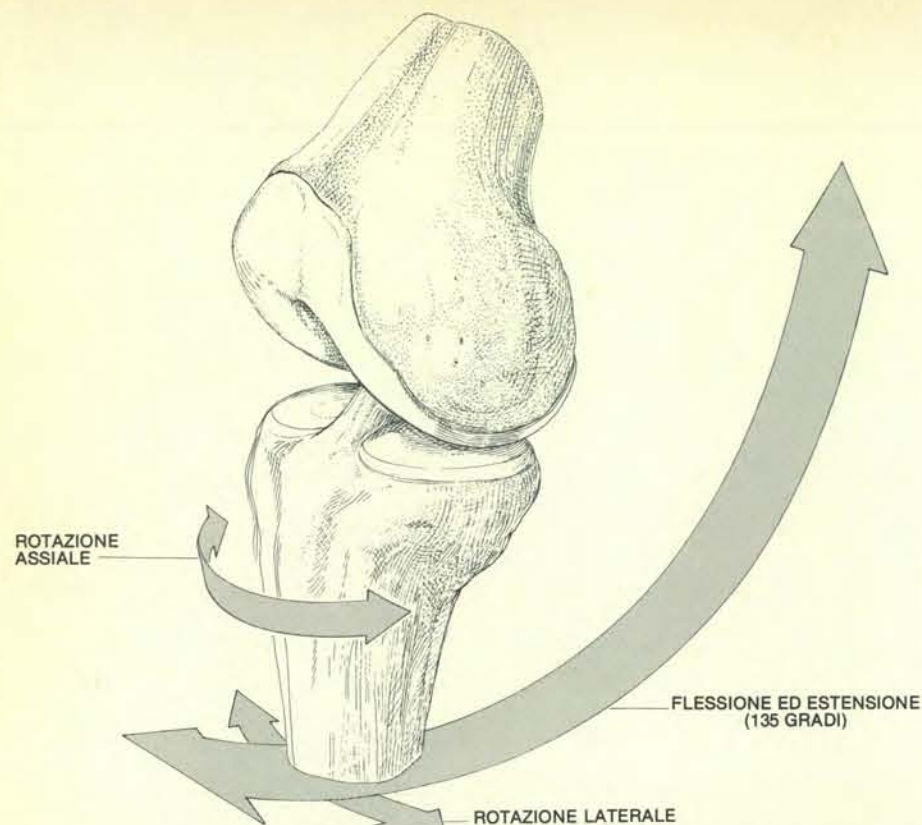
Il criterio di un movimento triassiale indicava che una sorta di giunto a sfera avrebbe costituito il mezzo migliore per ottenere una interconnessione meccanica tra i componenti mentre un meccanismo a guide e binari avrebbe definito la varietà e i limiti dei movimenti. Riuscimmo in entrambi i propositi, utilizzando una cavità femorale e due inserzioni tibiali progettate in modo che ogni parte della superficie articolare fosse rinforzata da inserti di metallo (*si veda la figura a pagina 37*). Il proposito di ottenere una decelerazione controllata dell'arto al punto di massima estensione del ginocchio è stato infine raggiunto mediante una collocazione appropriata della sfera e della cavità rispetto ai binari e alle guide, realizzati con una attenta scelta di raggi di curvatura variabili e fornendo quindi un'azione simile a quella di una camma.

In prove di carico effettuate su ossa prelevate da cadaveri, abbiamo trovato che sia la protesi di Walldius sia la protesi sferocentrica erano in grado di sostenere carichi maggiori del ginocchio nor-

**Il modo di funzionamento del ginocchio artificiale «sferocentrico» è mostrato in prospettiva frontale e laterale. Si tratta di una protesi articolata, con snodo articolare realizzato secondo lo schema del giunto a sfera. Le superfici articolari sono in metallo su materiale plastico.**

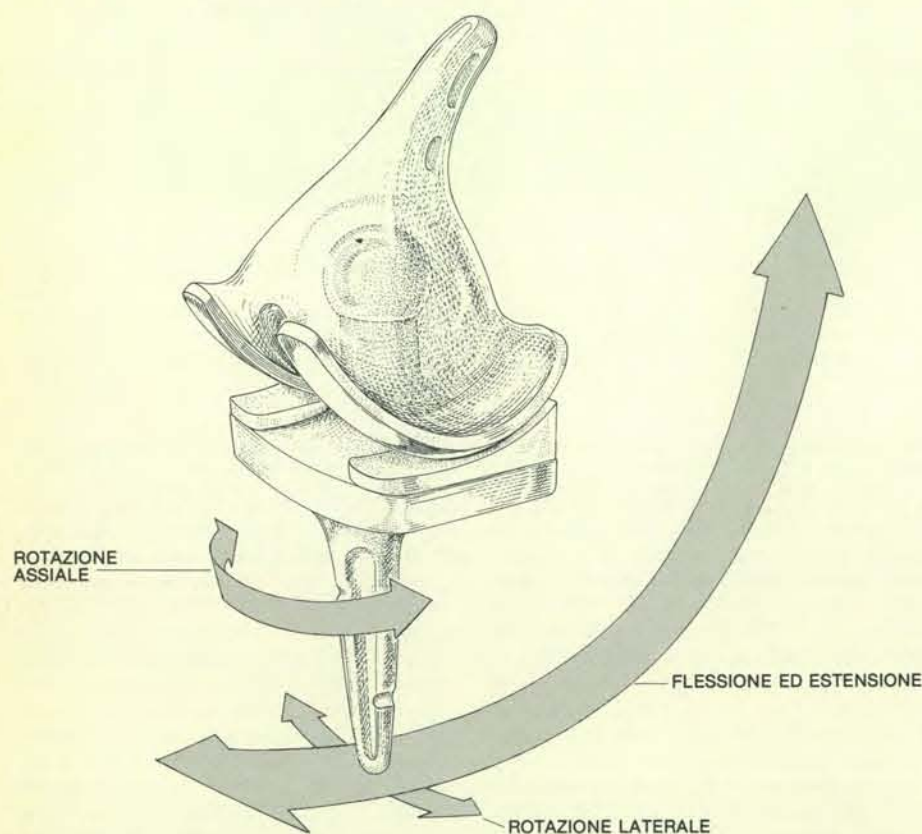
ROTAZIONE ASSIALE

FLESSIONE ED ESTENSIONE (135 GRADI)

ROTAZIONE LATERALE

**Raffigurazione schematica del movimento triassiale di cui è capace il ginocchio normale. Oltre a flettersi di 135 gradi, il ginocchio può ruotare assialmente, come quando si ruota il piede da un lato all'altro, facendo perno sulla caviglia, e lateralmente, il che significa che la tibia ruota rispetto al femore. Una simile rotazione si realizza quando la gamba viene flessa e quindi estesa.**

ROTAZIONE ASSIALE

FLESSIONE ED ESTENSIONE

ROTAZIONE LATERALE

**Raffigurazione della gamma di movimenti permessa dal ginocchio sferocentrico. Teoricamente un paziente dotato di questo tipo di protesi può ottenere una flessione e una estensione di 120 gradi, una rotazione assiale di 30 gradi e una rotazione laterale di 5 gradi. Il movimento è controllato soprattutto da guide e binari, che formano una serie di raggi di curvatura variabili.**

male, prima di cedere. Di particolare interesse furono le misure di carico dinamico di ciascuna protesi, relative all'arresto del movimento al termine della estensione. Il carico poteva essere misurato sotto forma di energia accumulata.

Abbiamo provato diverse versioni della protesi sferocentrica nei nostri laboratori, allo scopo di avvicinarci il più possibile nella protesi alle caratteristiche di accumulazione di energia del ginocchio normale. La versione definitiva del ginocchio presenta carichi di accumulazione di energia che sono uguali a un terzo di quelli sperimentati con il ginocchio di Walldius, che possiede un sistema di arresto in metallo contro metallo. Dal momento che i carichi coinvolti nell'arresto del movimento di estensione vengono trasmessi alle zone in cui la protesi è fissata all'osso per mezzo di cemento, sembra verosimile che la protesi sferocentrica incontrerà minori problemi di distacco, rispetto al ginocchio di Walldius e alle protesi con articolazione a ginglimo fisso con vincoli analoghi.

Nel 1973 il ginocchio sferocentrico era pronto ad affrontare il giudizio clinico. I nostri criteri per la scelta dei pazienti erano (e lo sono tutt'ora) la presenza di una grave deformazione, di una grande instabilità o di un cedimento della protesi precedente. La prima prova coinvolse 22 pazienti e 25 interventi. Abbiamo seguito questi pazienti dal momento dell'inserimento del ginocchio artificiale e abbiamo constatato come siano andati incontro a una notevole riduzione del dolore e fossero in grado di svolgere una attività molto maggiore.

Fino a oggi, al Medical Center dell'Università del Michigan sono stati impiantati più di 120 ginocchi del tipo sferocentrico. Ciononostante, la protesi deve essere ancora classificata come sperimentale. Sembra che le caratteristiche del movimento triassiale e della decelerazione controllata abbiano effettivamente ridotto il problema del distacco dei componenti, ma solo ulteriori esperienze potranno confermare se il distacco è evitabile sui lunghi periodi.

Una recente indagine condotta nell'ambito del Rehabilitation Engineering Program presso la Northwestern University, ha dimostrato che nel 1976, negli Stati Uniti, sono state portate a termine 80 000 sostituzioni totali dell'anca e 30 000 sostituzioni del ginocchio. I chirurghi, rispondendo al questionario che era parte integrante della indagine, hanno fatto presente che, se fossero state disponibili protesi più affidabili, avrebbero potuto eseguire un numero di interventi 1,12 volte superiore, per quanto riguarda la sostituzione dell'anca, e 1,76 volte superiore, per quanto riguarda la sostituzione del ginocchio. I dati suggeriscono che sono necessari ulteriori ricerche e perfezionamenti prima che la sostituzione del ginocchio umano possa essere considerata qualcosa di più di un successo qualificato.

Secondo il nostro punto di vista, si deve progredire in due direzioni. La prima ha a che fare con la fase di progettazione del ginocchio artificiale. Si stanno approntando materiali da impianto nuovi, polimerici, metallici e ceramici. Essi dovrebbero migliorare la prestazione meccanica e fornire in fase di progettazione maggiore possibilità di variazione. Il proseguimento delle ricerche sui carichi sia del ginocchio normale, sia di quello malato, consentirà di controllare meglio in laboratorio i nuovi progetti.

La seconda direzione in cui è necessario progredire riguarda la fissazione delle protesi all'osso. Il distacco costituisce la causa principale di cedimento del ginocchio artificiale (più o meno il 10 per cento in due anni). Finora non si sa abbastanza sul fenomeno per essere in grado di prevedere quali pazienti potrebbero andare incontro a una simile evenienza. Sicuramente la frequenza e la gravità del distacco aumentano con il passare del tempo e si può affermare che il fenomeno si verifica molto probabilmente in pazienti attivi o di maggior peso corporeo, con richiesta quindi alla protesi di prestazioni rilevanti.

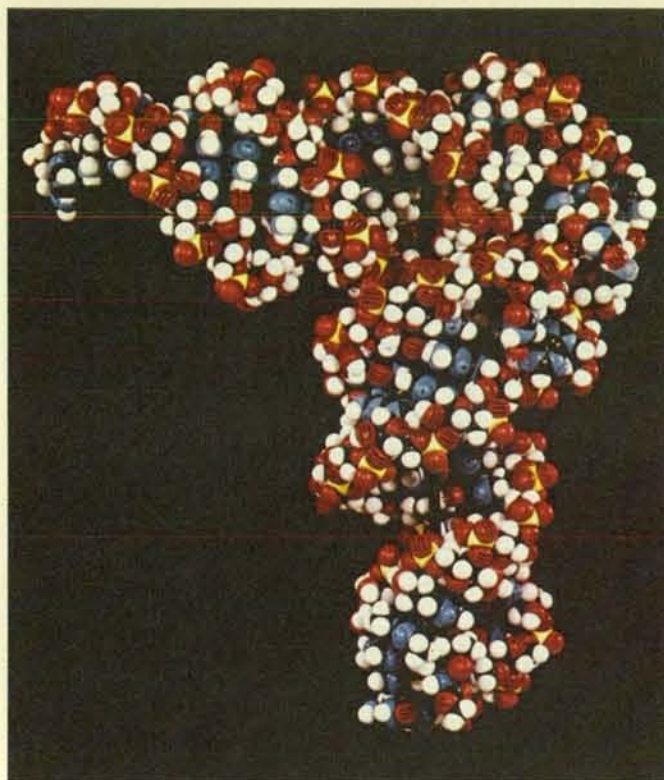
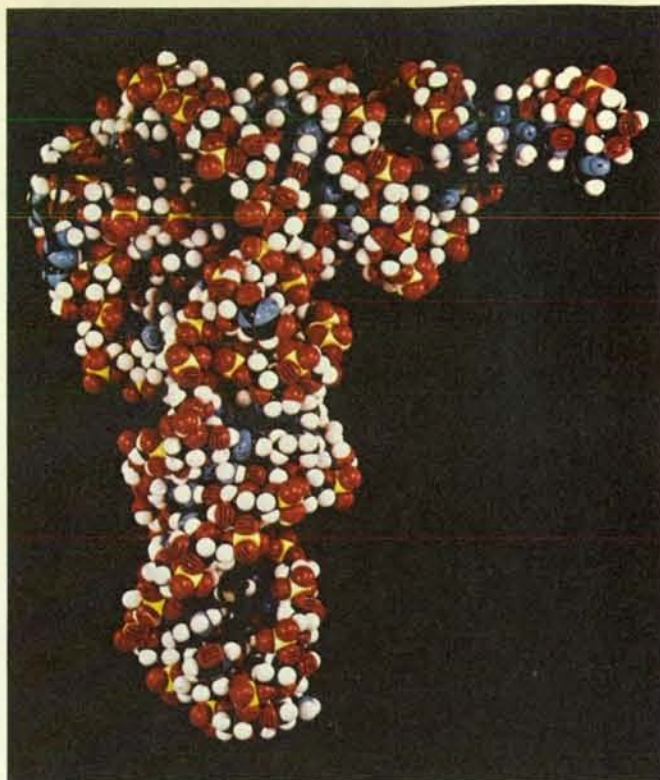
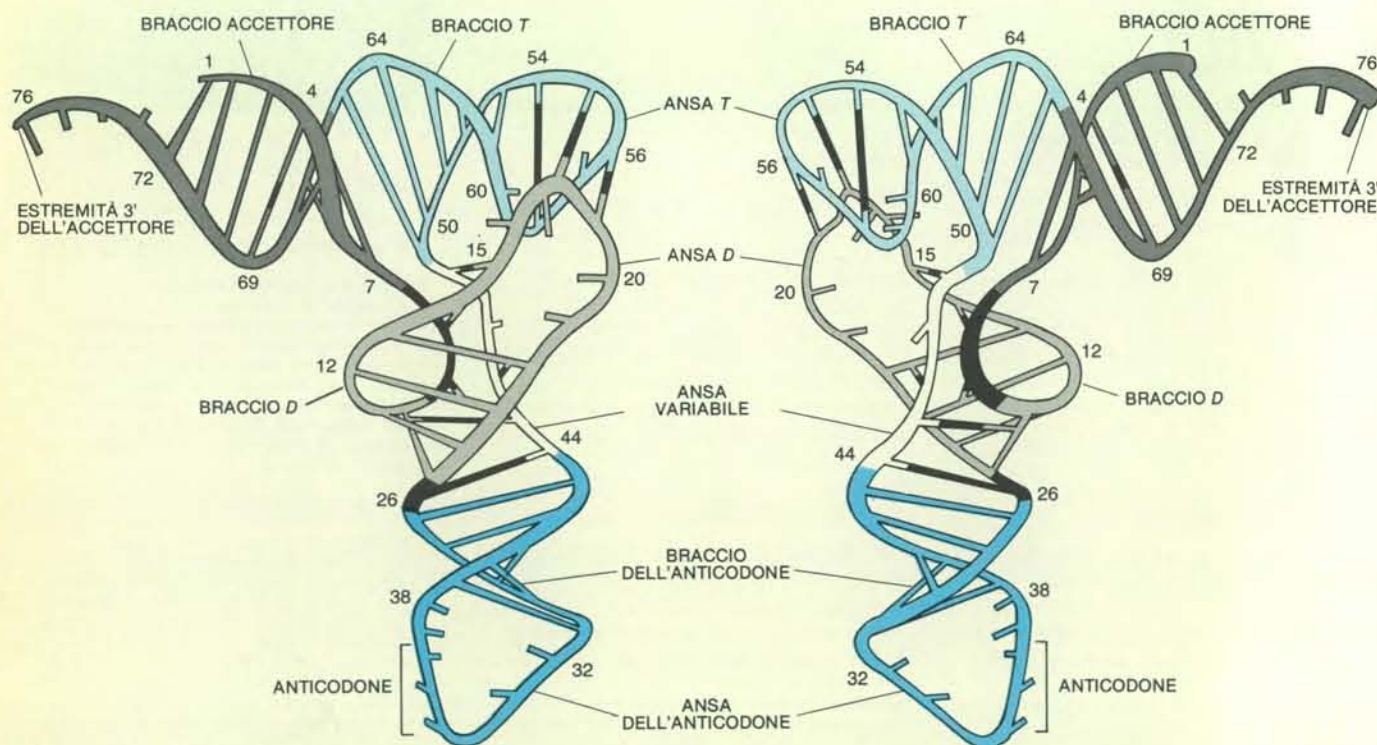
Abbiamo continuato a studiare questo problema nei nostri laboratori, imponendo carichi sulle zone di contatto tra osso

**In queste due fotografie sono mostrati tre componenti della protesi sferocentrica, a sinistra montati e a destra smontati. Le superfici articolari metalliche (la sfera e i binari), sono realizzate in una lega finemente levigata di cobalto, cromo e molibdeno, e le superfici articolari in materiale plastico (la cavità e le guide) sono realizzate in polietilene ad alto peso molecolare. Per impiantare la protesi il chirurgo inserisce prima il componente femorale e quindi quello tibiale. La cavità in polietilene viene sistemata sopra la sfera e quindi sospinta nel componente femorale, dove un meccanismo di arresto la tiene in sede. Una caratteristica del ginocchio sferocentrico è che tutte le superfici articolari in materiale plastico sono rinforzate da metallo. Nuovi materiali da impianto, come una migliore conoscenza delle cause di cedimento delle protesi, consentiranno ulteriori miglioramenti nella sostituzione chirurgica dell'articolazione del ginocchio.**

e cemento. I risultati indicano che il polimetilmetacrilato si fissa meglio all'osso spugnoso (il tessuto osseo situato presso l'estremità delle ossa lunghe) che non all'osso corticale (il tessuto osseo più compatto delle estremità). I risultati dimostrano anche ampie variazioni nelle proprietà dell'osso spugnoso in individui diversi; in ossa diverse del medesimo individuo e anche in zone diverse del medesimo osso.

Questi studi sono stati condotti su ossa prelevate da cadaveri, mentre il tessuto osseo vivente è un tessuto attivo che risponde ai carichi opponendo resistenza o rimodellandosi. Noi e moltri altri ricercatori riteniamo che il distacco della protesi sia da porre in relazione con questo meccanismo. In seguito all'impianto di una protesi l'osso a contatto con il cemento si riassorbe e si rimodella per riadattarsi ai nuovi carichi. Si stabilisce allora un equilibrio precario, in cui a volte prevale il riassorbimento e si ha il distacco della protesi. Solo esperimenti su animali vivi, accuratamente progettati, possono consentire di conoscere a fondo il problema di cosa avviene tra osso e cemento.

Altri ricercatori stanno studiando due sistemi alternativi di fissazione. Uno di questi sfrutta le proprietà piezoelettriche dell'osso. I potenziali elettrici generati dalla imposizione di carichi sull'osso, sia naturali, sia imposti dall'esterno, possono stimolare la crescita dell'osso e quindi favorire la fissazione della protesi. Il secondo metodo consiste nell'impiantare protesi dotate di superfici porose che lasciano spazio alla crescita dell'osso, in modo che la protesi si fissa direttamente. Matrici realizzate in materiale ceramico, polimerico e metallico, sagomate in modo opportuno, hanno dato in questo senso risultati promettenti.

La sostituzione del ginocchio umano rappresenta un campo di sperimentazione molto attivo e in rapida evoluzione. Le 30 000 sostituzioni effettuate nel 1976, soprattutto in pazienti già resi invalidi dalla artrite in modo cronico, hanno permesso la riduzione del dolore e il miglioramento funzionale nella maggior parte dei casi. Si prevedono in questo campo ulteriori perfezionamenti. Tra questi, il miglioramento dei materiali per la realizzazione delle protesi, l'identificazione di tecniche di laboratorio adatte alla verifica dei progetti e migliori prove di deambulazione in laboratorio per poter valutare le condizioni del paziente prima e dopo l'operazione. Ciononostante, la causa principale del fallimento clinico, cioè la tendenza al distacco dei componenti della protesi, deve essere eliminata prima che la sostituzione chirurgica dell'articolazione del ginocchio si estenda. Anche allora però, i pazienti, i chirurghi e i tecnici non dovranno dimenticare che l'articolazione naturale del ginocchio umano costituisce un ottimo risultato dell'ingegneria evolutiva. Sebbene possa essere imitata con crescente successo, l'articolazione naturale non potrà mai essere copiata alla perfezione.

Modello compatto a sfere di t-RNA-fenilalanina di lievito. Esso ha una struttura molto simile a quella reale della molecola ed è stato costruito sulla base di analisi diffrattometriche mediante i raggi X, condotte nei laboratori degli autori al Massachusetts Institute of Technology e alla Duke University School of Medicine. La catena polinucleotidica del t-RNA risulta ripiegata in una struttura compatta a *L*. Durante la sintesi proteica, la fenilalanina si attacca all'estremità del braccio orizzontale della *L*. Le tre basi nucleotidiche all'estremità del braccio verticale riconoscono il codice genetico per la fenilalanina sul filamento di RNA messaggero e l'amminoacido viene trasferito alla catena polipeptidica in formazione. Nel modello le sfere di carbonio sono nere, quelle di ossigeno rosse, di azoto azzurre, di fosforo gialle e di idrogeno bianche.

Schema del tipo di ripiegatura subita dalla catena polinucleotidica del t-RNA-fenilalanina di lievito. Lo scheletro della molecola, formato da ribosio e fosfato, è disegnato come un tubo avvolto a elica, mentre le sbarre trasversali rappresentano le coppie di basi nucleotidiche presenti nelle regioni degli steli. I monconi rappresentano quelle basi che non sono interessate nella formazione di legami a idrogeno da base a base. Le differenti sfumature di colore rispecchiano quelle che compaiono nel modello a trifoglio della pagina a fronte.



# La struttura tridimensionale dell'RNA di trasporto

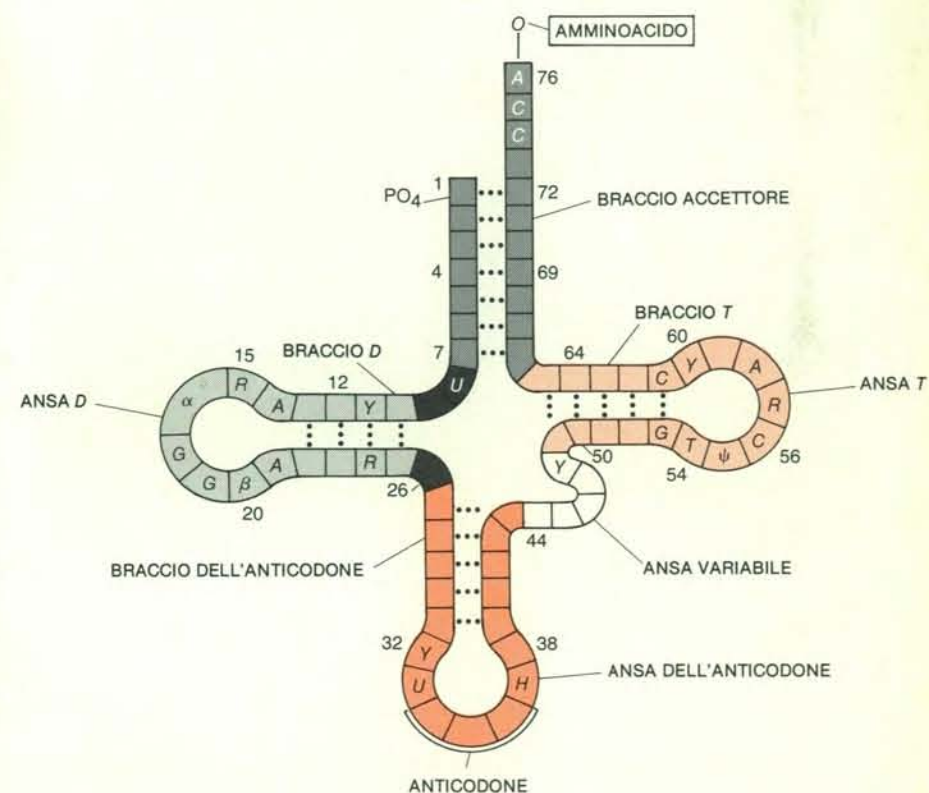
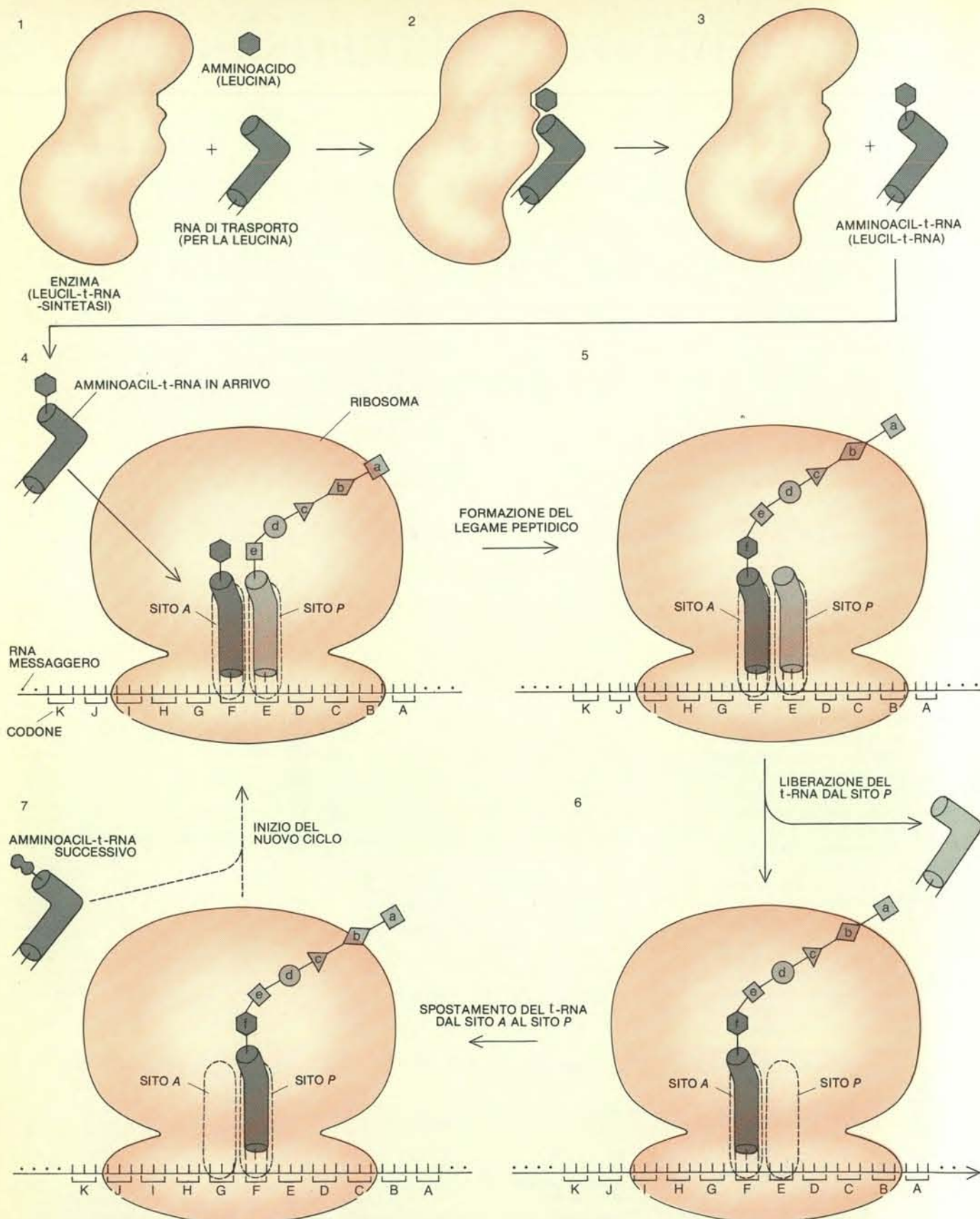
*La conoscenza della struttura tridimensionale di questo acido nucleico ha ora chiarito il meccanismo con il quale avviene la traduzione del codice genetico nella sequenza di amminoacidi di una proteina*

di Alexander Rich e Sung Hou Kim

È ampiamente riconosciuto oggi che le istruzioni per l'«assemblaggio» e l'organizzazione di un sistema vivente sono incorporate nelle molecole di DNA, contenute nella cellula. La sequenza di basi nucleotidiche lungo la catena lineare di tali molecole specifica la struttura di quelle migliaia di proteine che sono i materiali costitutivi della cellula e i catalizzatori delle sue complesse reazioni biochimiche. Di per sé, tuttavia, una molecola di DNA è paragonabile piuttosto a un nastro registratore magnetico: l'informazione che contiene nella sua struttura non può venire espressa senza un meccanismo di decodificazione.

Lo sviluppo di questo meccanismo è stato uno degli eventi fondamentali verificatisi agli inizi della vita, circa quattro miliardi di anni fa. Gradatamente si è evoluto un sistema biochimico di base, in cui la sequenza di nucleotidi del DNA viene dapprima trascritta nella sequenza complementare dell'RNA messaggero (abbreviato m-RNA). L'RNA messaggero dirige quindi l'assemblaggio degli amminoacidi nella sequenza lineare specifica, caratteristica di una data proteina: questo processo si chiama traduzione.

Una funzione cardine nella traduzione è svolta da un altro tipo di RNA: l'RNA di trasporto (abbreviato t-RNA). Le molecole di questo acido nucleico formano una classe di piccole catene polinucleotidiche globulari (distinte dalle catene polinucleotidiche filamentose, quali sono quelle del DNA e dell'm-RNA), lunghe da 75 a 90 nucleotidi, che fungono da vettori per il trasporto degli amminoacidi dallo stato libero all'interno della cellula allo stato legato nella catena proteica. Si tratta di una funzione di vitale importanza in quanto agisce come intermediaria tra il linguaggio in termini di acido nucleico del codice genetico e il linguaggio in termini di amminoacidi della cellula durante lo svolgimento delle sue funzioni; essa ha reso l'RNA di trasporto un soggetto del massimo interesse in biologia molecolare. Di recente, in una fase importante della ricerca che mira a raggiungere la comprensione del processo di traduzione in termini molecolari precisi, si è ricostruita con un elevato grado di risoluzione, la struttura tridimensionale di una molecola di t-RNA.

La traduzione della sequenza nucleotidica dell'RNA messaggero in proteina procede attraverso due tappe principali. Innanzitutto, una molecola di amminoacido si attacca a una particolare moleco-

Il modello a trifoglio rappresenta il tipo di struttura ripiegata bidimensionale della molecola dell'RNA di trasporto (o t-RNA), ricavato per la prima volta nel 1965 a partire dalla sequenza delle unità nucleotidiche costitutive del complesso t-RNA-alanina di lievito. In seguito si è trovato che tale modello è valido anche per le sequenze nucleotidiche di circa 100 t-RNA isolati da cellule vegetali, animali e batteriche. Sono qui indicate le basi nucleotidiche che si trovano nelle stesse posizioni in tutte le sequenze di t-RNA. I bracci o steli del trifoglio sono costituiti da basi complementari, presenti in differenti parti della catena polinucleotidica e che si uniscono mediante legami a idrogeno, determinando così un ripiegamento su se stessa della catena. Il numero di nucleotidi nei bracci e anse ad anello è generalmente costante, tranne che in due zone dell'ansa *D* designate con $\alpha$ e $\beta$ (le quali consistono nei vari RNA di un numero variabile di nucleotidi da uno a tre) e nell'ansa variabile (che ha 4 o 5 nucleotidi, ma può anche averne 21). Le abbreviazioni indicano: *A*, adenosina; *G*, guanosina; *C*, citidina; *U*, uridina; *R*, adenosina o guanosina; *Y*, citidina o uridina; *T*, ribotimidina; $\Psi$, pseudouridina; *H*, adenosina o guanosina modificate.

La funzione dell'RNA di trasporto nella sintesi di una molecola proteica consiste nel produrre una catena di amminoacidi che rispecchia la sequenza nucleotidica dello stampo rappresentato dall'RNA messaggero. Per cominciare, una grossa molecola enzimatica, chiamata amminoacil-t-RNA-sintetasi, attacca con legame covalente un amminoacido alla molecola di RNA di trasporto specifica per esso (*1-3*). Questo t-RNA con l'amminoacido attaccato si lega a sua volta al sito *A* del ribosoma, cioè dell'organello in cui avviene il montaggio degli amminoacidi nella catena polipeptidica di una proteina. Questa interazione richiede la formazione di legami a idrogeno specifici tra le tre basi del codone sul filamento di RNA messaggero, che specifica per un amminoacido, e le tre basi dell'anticodone dell'RNA di trasporto (*4*). Una molecola di RNA di trasporto, presente nel sito adiacente *P*, trasferisce a questo punto la catena polipeptidica che si sta formando all'RNA di trasporto che si trova nel sito *A* (*5*). Il t-RNA che rimane «vacante» abbandona il sito *P* e il ribosoma si sposta lungo l'RNA messaggero della distanza di un codone, in modo che l'RNA di trasporto per la catena polipeptidica passi dal sito *A* al sito *P* (*6*, *7*).

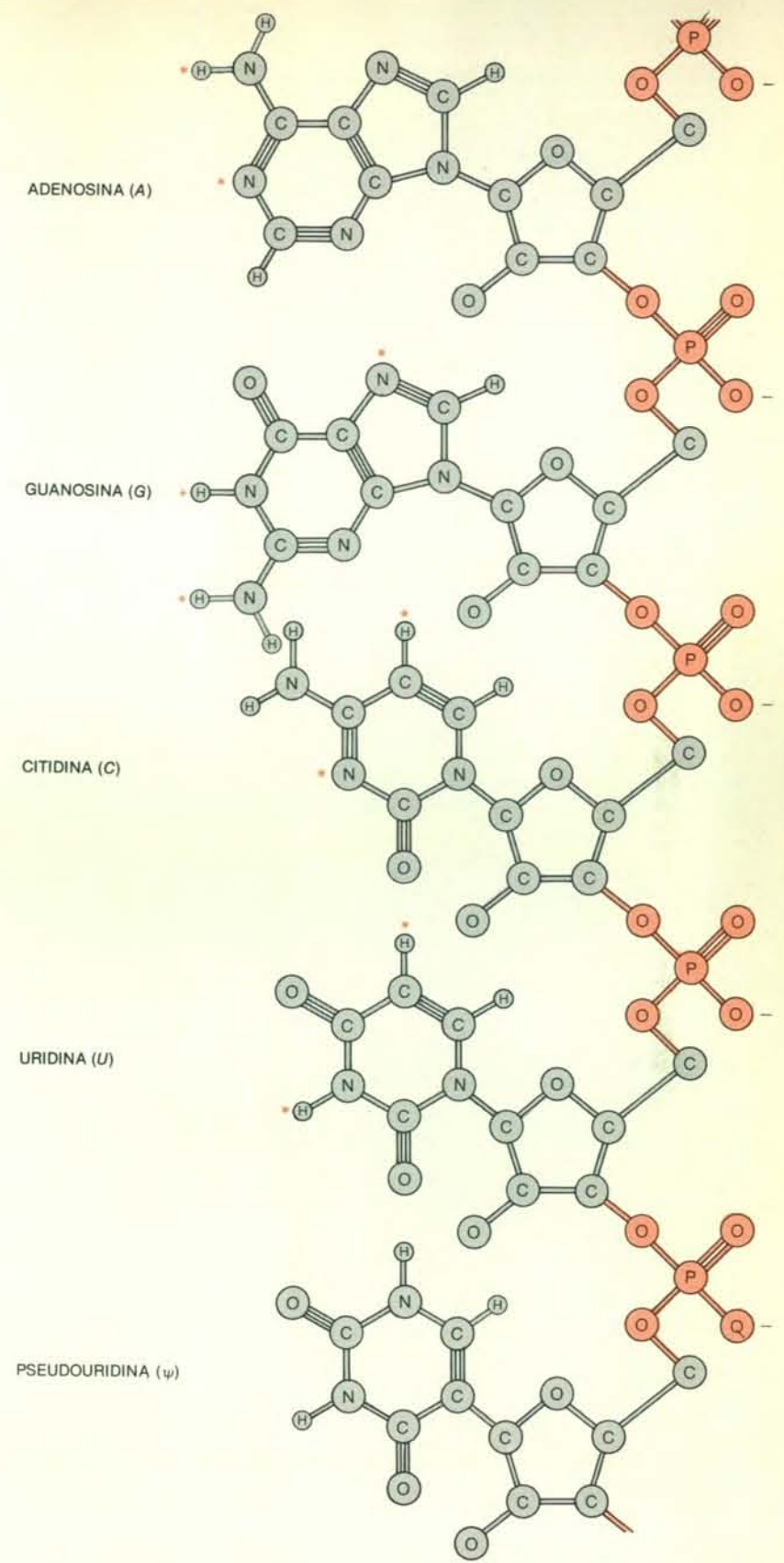


la di RNA di trasporto, reazione questa catalizzata da un enzima di notevoli dimensioni, una amminoacil-t-RNA-sintetasi. Nelle cellule esistono molti diversi tipi di sintetasi, ognuno specifico per uno dei venti diversi amminoacidi che si trovano nelle proteine. Per esempio, la leucil-t-RNA-sintetasi lega selettivamente se stessa sia all'amminoacido leucina, sia al t-RNA per la leucina: si forma, allora, e viene reso disponibile, un complesso costituito da leucina e da t-RNA per la leucina. Quando un t-RNA ha attaccato a sè un amminoacido, è pronto per partecipare alla seconda tappa del processo di sintesi proteica.

Questa seconda fase, il montaggio degli amminoacidi in una catena, si svolge in un organello cellulare noto come ribosoma: un aggregato di più di 50 diverse molecole proteiche e di tre molecole di RNA. Si tratta in effetti di un complesso pezzo del congegno molecolare, progettato per favorire la traduzione della sequenza polinucleotidica dell'RNA messaggero nella sequenza polipeptidica della proteina. Anche se non sono stati ancora messi a punto i particolari esatti di questo processo, sono però note le sue caratteristiche generali.

Ogni amminoacido in una proteina viene specificato da un gruppo di tre basi nucleotidiche adiacenti, designato come codone e presente sul filamento dell'RNA messaggero. In quest'ultimo sono presenti quattro tipi di basi nucleotidiche e, pertanto, è possibile un totale di $4^3$ o 64 codoni. La relazione tra i codoni e gli amminoacidi che essi specificano è il codice genetico: il fatto che questo risulti uguale in tutti gli organismi viventi è una prova indiscutibile dell'uniformità dell'organizzazione della vita a livello molecolare.

Nel ribosoma sono presenti due siti, interessati alla traduzione. Uno è il sito *A*, cioè il sito di legame per l'amminoacil-t-RNA. È in questa posizione che la molecola di RNA di trasporto con l'amminoacido legato si lega a sua volta al ribosoma. Essa si piazza proprio qui grazie, in parte, a una serie di interazioni specifiche con l'RNA messaggero, che si è già associato con il ribosoma. Tre basi nucleotidiche speciali, presenti nella molecola dell'RNA di trasporto e designate come anticodone, interagiscono con tre basi complementari di un codone presente nell'RNA messaggero. Questa interazione comporta la formazione di legami direzionali deboli, denominati legami a idrogeno, in cui un atomo di idrogeno con una debole carica positiva viene condiviso da due altri atomi, dotati di una debole carica negativa. I legami a idrogeno sono quelle forze che tengono anche unite le basi nucleotidiche complementari nella doppia elica del DNA: la base guanina su un filamento dell'elica è sempre appaiata con la base citosina sull'altro filamento dell'elica mentre la base adenina è appaiata con la base timina.

Nelle immediate adiacenze del sito *A*, nel ribosoma, si trova il sito di legame

I nucleosidi, formati da un base nucleotidica attaccata allo zucchero ribosio, si uniscono tra loro mediante gruppi fosfato ($PO_4$), dotati di carica negativa, e formano la catena polinucleotidica dell'RNA di trasporto. I quattro principali nucleosidi presenti nella molecola sono l'adenosina, la guanosina, la citidina e l'uridina. L'RNA di trasporto include anche molti nucleosidi modificati, una cinquantina dei quali sono stati identificati. La modificazione più comune consiste nella sostituzione di un atomo di idrogeno con gruppo metilico ($CH_3$), reazione che è catalizzata da enzimi speciali e che si verifica nei siti indicati da un asterisco. Si riscontrano anche altre modificazioni strutturali. Per esempio, il nucleoside pseudouridina (ψ) presenta la propria base attaccata allo zucchero ribosio attraverso un atomo di carbonio, invece che un atomo di azoto.

per il peptidil-t-RNA, o sito *P*. La molecola di RNA di trasporto, con la catena di amminoacidi in accrescimento, attaccata a essa, si lega in questo punto e interagisce in maniera specifica con la successiva tripletta di basi, o codone, presente sulla catena dell'RNA messaggero. Nel corso della sintesi proteica, la catena polipeptidica in accrescimento si stacca, nel sito *P*, dalla molecola di t-RNA e si trasferisce all'estremità del singolo amminoacido attaccato alla molecola di t-RNA che si trova nel sito *A*. Avvenuto questo passaggio (le reazioni di distacco e di ricongiungimento avvengono grazie a un enzima presente nel ribosoma), la catena polipeptidica in via di formazione si è allungata di un amminoacido. La molecola di t-RNA rimasta «vacante» si stacca allora dal sito *P* e l'RNA messaggero con il peptidil-t-RNA appena allungatosi si sposta dal sito *A* al sito *P*. A questo punto, un nuovo t-RNA con un amminoacido attaccato si fa strada e raggiunge il ribosoma, sistemandosi nel sito *A* rimasto libero, grazie all'interazione specifica tra le proprie basi, che costituiscono l'anticodone, e quelle del codone successivo, presente sul filamento di RNA messaggero. Il sistema è ora ritornato al punto di partenza, pronto per ricominciare un altro ciclo di eventi, in cui alla catena si aggiungerà un amminoacido in più. Questa graduale aggiunta si ripete fino a quando non è stata sintetizzata tutta la proteina.

Il processo di allungamento della catena polipeptidica è piuttosto rapido: si ripete fino a venti volte al secondo in una cellula batterica e circa una volta al secondo in una cellula di mammifero. Per esempio, la molecola dell'emoglobina è una grossa proteina che consiste di quattro catene polipeptidiche con circa 140 amminoacidi ciascuna. La sintesi di una di queste catene impiegherebbe sette secondi in una cellula batterica e due o tre minuti in una cellula di mammifero. Pur essendo questa velocità di sintesi abbastanza elevata, si notano sorprendentemente pochi errori nella traduzione, dato che il congegno del ribosoma assicura una precisa corrispondenza tra ogni molecola di RNA di trasporto e l'RNA messaggero. Il processo è inoltre efficiente, poiché vi sono di solito parecchi ribosomi all'opera, nel tradurre un singolo filamento di RNA messaggero.

Per capire come l'RNA di trasporto porti un amminoacido nel ribosoma, trasferendolo poi alla catena polipeptidica in formazione, è essenziale conoscere la struttura della sua molecola. Uno dei primi indizi riguardanti tale struttura è emerso dalla sequenza nucleotidica di un t-RNA di lievito, specifico per l'amminoacido alanina, sequenza che è stata determinata nel 1965 da Robert H. Holley e dai suoi collaboratori della Cornell University. Questi ricercatori hanno notato l'esistenza di certe regioni della sequenza che sarebbero state complementari se la catena si fosse ripiegata su se stessa. In particolare, queste regioni potevano formare tra loro legami a idrogeno, proprio come nella doppia elica del DNA avviene l'appaiamento delle basi (negli RNA, però, la base adenina si appaia con l'uracile invece che con la timina). La catena polinucleotidica dell'RNA di trasporto poteva così essere disposta in modo da contenere regioni a doppio filamento, con legami a idrogeno, dette steli o braccia, e regioni senza legami a idrogeno, chiamate anse ad anello. L'insieme di steli e anse, che si postulava, assomigliava a un trifoglio e, pertanto, la struttura prese il nome di «struttura a trifoglio».

Una caratteristica della sequenza nucleotidica dell'RNA di trasporto è la presenza in essa di molte basi insolite, la maggior parte delle quali sono le comuni basi dell'RNA, modificatesi per aggiunta di uno o più gruppi metilici ($CH_3$). In conseguenza di questo fatto, alcune parti della struttura a trifoglio sono state chiamate in base alle basi modificate presenti in esse. Per esempio, l'ansa *T* viene così chiamata perché contiene la timina (*T*), che si trova nel DNA, ma non nell'RNA a eccezione dell'RNA di trasporto. Alla stessa stregua, l'ansa *D* include, di solito, la base modificata diidrouracile (*D*). Altre regioni del trifoglio includono l'ansa variabile, che nei diversi t-RNA ha un numero diverso di nucleotidi (da quattro a ventuno), l'ansa dell'anticodone, che include le tre basi dell'anticodone, e lo stelo, o braccio, accettore, che si fissa all'amminoacido specifico per quel particolare t-RNA.

Un aspetto tipico della struttura a trifoglio è la presenza di sequenze nucleotidiche costanti in tutte le cento sequenze di t-RNA che sono state finora determinate. Anche il numero di coppie di basi nelle regioni a stelo è costante: sette nel braccio accettore, cinque in quello *T*, cinque in quello dell'anticodone e tre o quattro nel braccio *D*, caratteristiche che si conservano nelle molecole di t-RNA dalle piante agli animali, ai batteri e ai virus. In effetti, la disposizione delle braccia, delle anse ad anello e dei nucleotidi costanti, che si riscontra nella struttura a trifoglio, risulta avere la stessa universalità del codice genetico. Questa costanza è stata in buona parte spiegata in seguito con la definizione della struttura tridimensionale del t-RNA.

Oggi la struttura tridimensionale delle grosse molecole biologiche viene determinata di solito mediante analisi diffrattometrica ai raggi X dei cristalli molecolari. Un cristallo molecolare è un aggregato di molecole, impacchettate in

**L'analisi diffrattometrica mediante raggi X è stata uno degli strumenti utilizzati dagli autori per ricostruire la struttura tridimensionale dell'RNA di trasporto. Questo diffrattogramma è stato ottenuto dirigendo raggi X su un reticolo cristallino costituito da molecole di RNA di trasporto e registrando poi le radiazioni diffratte su una pellicola fotografica. La distribuzione delle macchie fornisce informazioni sulla disposizione degli elettroni all'interno del cristallo.**



**Il modello dettagliato dello «scheletro» della molecola di t-RNA legato all'alanina di lievito mostra le interazioni tra basi nucleotidiche, costituite da legami a idrogeno. È stato ottenuto nel 1974 partendo da uno studio cristallografico con i raggi X a una risoluzione di 0,3 nanometri. Questa ricostruzione è stata elaborata al calcolatore da S.H. Kim. Il ribosio e il gruppo fosfato sono in colore, le basi in grigio.**

modo da avere una disposizione tridimensionale regolare. Quando raggi X, con una lunghezza d'onda paragonabile alla distanza che esiste tra gli atomi, vengono diretti in modo da penetrare in un cristallo, sono diffratti, o deviati, in una varietà di direzioni dalle nubi elettroniche degli atomi presenti nel reticolo cristallino. Il diffrattogramma appare come una serie di macchie su una lastra fotografica per raggi X, in cui l'annerimento dell'emulsione è proporzionale all'intensità di ogni fascio deviato.

Questo diffrattogramma contiene una massa notevole di informazioni sulla struttura del cristallo. In primo luogo, l'ampiezza dell'onda deviata da un atomo è proporzionale al numero di elettroni che sono presenti in esso, per cui un atomo di carbonio devierà i raggi X con una intensità sei volte superiore a quella dell'atomo di idrogeno. In secondo luogo, le onde diffratte si ricombinano all'interno del reticolo cristallino dove si sommano o si annullano in relazione al fatto di essere o non essere in fase. Il modo in cui avviene questa ricombinazione dipende soltanto dalla disposizione degli atomi nel cristallo, per cui è possibile ricostruire l'immagine di una molecola dalla sua immagine di diffrazione dei raggi X.

Per analizzare la struttura tridimensionale di una grossa molecola di proteina o di acido nucleico, si prepara innanzitutto un cristallo della sostanza. Questo viene quindi introdotto in un tubo capillare e messo nella esatta posizione rispetto al fascio di raggi X e alla pellicola fotografica. Infine, viene fatto ruotare lungo ciascuno dei suoi assi, così da ottenere una serie di fotografie ai raggi X, in cui si nota una disposizione regolare di macchie di varie intensità. Ogni fotografia è, in realtà, una sezione bidimensionale ottenuta attraverso una disposizione tridimensionale di macchie.

Nella fase successiva, si misurano le intensità di tutte le macchie presenti sui diffrattogrammi o direttamente sulla pellicola o ricorrendo a un contatore Geiger. Sono necessarie però ulteriori informazioni prima di riuscire a stabilire la struttura tridimensionale, in particolare occorre conoscere le fasi dei fasci diffratti di raggi X, riguardo a un punto arbitrario, fissato nel cristallo. Ciò si ottiene inserendo nel reticolo cristallino come marcatori atomi di metalli pesanti come platino od oro. La loro aggiunta modifica leggermente il diffrattogramma e permette di calcolare le fasi dei fasci diffratti.

Avendo a disposizione questo dato, si può calcolare la densità elettronica in corrispondenza di un numero elevato di punti regolarmente spaziati nel cristallo. E ciò ricorrendo a una serie di Fourier: una somma infinita di seni e di coseni. Bisogna disporre di un calcolatore ad alta velocità per poter maneggiare l'enorme quantità di termini (superiore al miliardo), implicati nella determinazione della struttura di una grossa molecola di proteina o di acido nucleico. La prima molecola di questo tipo, la cui struttura è stata determinata in questo modo, è stata, nel 1958, la mioglobina. Oggi la tecnica viene praticamente utilizzata come routine nell'analisi strutturale delle grosse molecole.

Il prodotto finale che si ottiene è una mappa tridimensionale, che mostra la distribuzione degli elettroni nel cristallo. Generalmente essa è formata da una serie di sezioni parallele poste l'una sull'altra. Ogni sezione consiste di un foglio di plastica trasparente su cui la distribuzione della densità elettronica è rappresentata da linee di contorno nere, che assomigliano a quelle di una mappa topografica. Il fattore critico nell'interpretazione di una mappa di densità elettronica è costituito dalla risoluzione, che è determinata dal numero dei punti d'informazione (cioè dei punti di cui si prende in considerazione l'intensità) che compaiono nella serie di Fourier. Per esempio, una mappa avente risoluzione di 0,6 nanometri, ricavata dalle macchie più interne del diffrattogramma, può rivelare la forma generica della molecola, ma pochi altri particolari strutturali (1 nanometro corrisponde a $10^{-9}$ metri ed equivale all'incirca al diametro di un atomo di idrogeno). Sono necessarie mappe con una risoluzione più elevata per delineare gruppi di atomi, che possono essere distanti tra loro 0,3-0,4 nanometri, oppure singoli atomi, distanti 0,1-0,2 nanometri. Generalmente una grossa molecola viene analizzata a diversi livelli di risoluzione, per mettere in evidenza vari aspetti della sua struttura. La massima risoluzione di una analisi con i raggi X è però determinata dal grado di perfezione del cristallo: per le grosse molecole biologiche, la miglior risoluzione che si riesce generalmente a ottenere è di circa 0,2 nanometri.

**I segmenti elicoidali della molecola di t-RNA, corrispondenti ai quattro bracci dello schema a pagina 44, vengono qui rappresentati da nastri. Le due regioni a elica hanno una disposizione ad angolo retto e forniscono così l'impalcatura strutturale per la ripiegatura a forma di *L*. Ogni regione consiste di circa 10 coppie di basi, ossia a un giro della doppia elica. Questa, nelle suddette regioni, è simile alla doppia elica del DNA, tranne che per il fatto che, nell'RNA di trasporto, i due filamenti sono formati da porzioni differenti della stessa catena polinucleotidica.**

Con l'RNA di trasporto, la prima tappa del processo - la cristallizzazione della molecola - è risultata essere un grosso ostacolo. Nel 1968, il nostro gruppo presso il Massachusetts Institute of Technology e altri ricercatori in cinque diversi laboratori hanno scoperto che si potevano cristallizzare diverse specie di t-RNA, sciogliendole in vari miscele di solventi che venivano poi fatte evaporare lentamente. Questa scoperta ha suscitato notevole entusiasmo tra i biologi molecolari, in quanto sembrava che permettesse di superare il maggior ostacolo e che la struttura tridimensionale dell'RNA di trasporto fosse ormai a portata di mano. Questo stato d'animo ha subito un duro contraccolpo quando ci siamo resi conto che, pur essendo state cristallizzate molte diverse specie di t-RNA, la maggior parte dei cristalli era disordinata. Pertanto, le immagini di diffrazione fornite da questi cristalli avevano una risoluzione molto bassa (generalmente tra i 1 e 2 nanometri) e potevano così rivelare molto poco sulla struttura particolareggiata della molecola. Scoprire che il t-RNA era cristallizzabile era tuttavia una bella soddisfazione, mitigata però dalla constatazione che era necessario altro lavoro prima di riuscire a ottenere il materiale adatto per l'analisi diffrattometrica.

In collaborazione con Gary J. Quigley e Fred L. Suddath, ci siamo dedicati al compito di stabilire le condizioni in cui il t-RNA formerebbe un cristallo ben ordinato, in grado di produrre un'immagine di diffrazione dei raggi X con un livello di risoluzione tale da rilevare la struttura tridimensionale della molecola. Per due anni abbiamo continuato a passare in rassegna moltissime specie di t-RNA e più diverse condizioni di cristallizzazione. Alla fine, abbiamo compiuto un'importante scoperta: l'aggiunta di spermina, una piccola molecola dotata di carica positiva, dava luogo alla formazione di un cristallo estremamente ordinato di un t-RNA, estratto da cellule di lievito e specifico per l'amminoacido fenilalanina. Questo cristallo, reso stabile dalla spermina, aveva un diffrattogramma la cui risoluzione era di quasi 0,2 nanometri.

Più tardi, nel 1972, in collaborazione con Alexander McPherson, Daryll Sneden, Jung-Ja Park Kim e Jon Weinzierl, abbiamo ottenuto una mappa di densità elettronica del cristallo che ci ha permesso di delineare lo «scheletro» della catena polinucleotidica del t-RNA con una risoluzione di 0,4 nanometri. A questo livello non era possibile percepire le singole basi della catena, ma i gruppi fosfato ($PO_4$) lungo lo scheletro della molecola, dotati di un'elevata densità elettronica, erano visibili come un filo di perle avvolto a elica nello spazio tridimensionale. Con nostra grande sorpresa, la catena polinucleotidica aveva un'organizzazione tale che la molecola risultava conformata a *L*, con un braccio della *L* costituito dallo stelo, o braccio, accettore e dallo stelo *T* e con l'altro braccio della *L* costituito dallo stelo D e dal braccio dell'anticodone. Le sequenze complementari, con i legami a idrogeno, che erano state identificate nella struttura a quadrifoglio, risultavano chiaramente visibili come doppie eliche di RNA. Le varie anse ad anello occupavano posizioni strategiche o a una estremità della molecola, o in corrispondenza del suo angolo, laddove le anse *T* e *D* risultavano avvolte insieme a elica, in maniera complessa.

**La formazione di legami a idrogeno tra basi complementari nelle regioni con avvolgimento a elica della molecola di RNA di trasporto segue lo schema proposto per la prima volta per la doppia elica del DNA da Watson e Crick (tranne che, nel t-RNA, l'uracile sostituisce la timina). L'adenina e l'uracile si appaiano con due legami a idrogeno, la guanina e la citosina con tre.**

Questo ripiegamento della molecola risultava del tutto inatteso. Nei pochi anni precedenti, un certo numero di ricercatori aveva riconosciuto le caratteristiche comuni a tutti i trifogli degli RNA di trasporto e aveva tentato di prevedere il tipo di ripiegamento della loro molecola. Tuttavia, come spesso accade, la natura si è rivelata più abile di quanto non si fosse immaginato. Il ripiegamento a forma di *L* spiega infatti un certo numero di osservazioni chimiche che erano andate accumulandosi.

Verso la metà del 1974, in collaborazione con Joel L. Sussman, Andrew H.-J. Wang e Nadrian C. Seeman, siamo riusciti a interpretare la mappa di densità elettronica con una risoluzione di 0,3 nanometri. La forma complessiva della molecola era la stessa di quella che si vedeva con una risoluzione di 0,4 nanometri, ma nel primo caso erano visibili molti più dettagli, tra cui le posizioni della maggior parte delle basi nucleotidiche. Quasi contemporaneamente, Jon Robertus, Brian F.C. Clark, Aaron Klug e collaboratori, presso il Laboratory of Molecular Biology del British Medical Research Council, a Cambridge, hanno descritto la loro analisi cristallografica ai raggi X di un RNA di trasporto con una risoluzione di 0,3 nanometri. Il loro materiale era lo stesso t-RNA legato alla fenilalanina, estratto da lievito e reso stabile dalla spermina, ma si trovava in una diversa forma cristallina. Tuttavia, anche se la molecola era «stipata» in un modo differente nel reticolo cristallino, il confronto tra le due strutture tridimensionali risultanti dalle analisi ha messo in evidenza che esse erano virtualmente identiche. Questa concordanza tra i risultati ottenuti dai due gruppi ha confermato che la struttura della molecola di t-RNA è indipendente dal modo in cui essa risulta stipata nel cristallo.

La mappa relativa alla molecola di t-RNA, con una risoluzione di 0,3 nanometri, ha confermato il nostro dato precedente, secondo cui essa risulta organizzata in due pile di basi nucleotidiche, disposte ad angolo retto l'una rispetto

**Interazioni insolite tra le basi rendono stabile la ripiegatura della molecola di RNA di trasporto. Nel braccio accettore, le basi guanina e uracile, normalmente non complementari, sono tenute unite da due legami a idrogeno come risultato di una lieve oscillazione laterale, o spostamento, in corrispondenza di una delle basi (*a*). Nell'ansa *T*, la 1-metiladenina si appaia con la timina, una forma modificata di uracile che ha un gruppo metilico in più (*b*). Nella regione centrale della molecola, immediatamente sotto l'angolo, la guanina e la citosina risultano appaiate, ma con due legami a idrogeno al posto dei normali tre. Questo appaiamento è di tipo *trans*, in quanto i residui del ribosio vengono a trovarsi sui lati opposti rispetto alla coppia di basi (*c*). Sempre nella regione centrale vi sono due sistemi di legame a idrogeno, che interessano tre basi nello stesso piano (*d*, *e*). Nella regione, invece, che congiunge il braccio *D* con il braccio dell'anticodone una guanina dimetilata si appaia con una adenina grazie a due legami a idrogeno (*f*). Questa coppia non può essere complanare a causa dei voluminosi gruppi metilici della guanina; le due basi formano tra loro un angolo di 25 gradi, come le pale di un propulsore. La dimetilguanina è localizzata in fondo al braccio *D* e l'adenina alla sommità del braccio dell'anticodone, una disposizione che rende stabile l'unione tra i bracci.**

all'altra. Queste pile presentano regioni con o senza avvolgimento a elica, corrispondenti rispettivamente agli steli o bracci e alle anse ad anello dello schema a trifoglio. La mappa con risoluzione elevata ha anche rilevato che le due regioni a elica consistono ciascuna di circa 10 coppie di basi, corrispondenti a un giro della doppia elica, e possiedono lo stesso tipo di legami a idrogeno tra basi nucleotidiche complementari che si trova nella doppia elica del DNA.

Nelle parti del t-RNA, che non presentano avvolgimento a elica, molte basi nucleotidiche sono orientate in modo che i loro rispettivi gruppi, tra cui si formano i legami a idrogeno, puntino verso l'interno della molecola, dove essi partecipano a una varietà di interazioni insolite, basate sempre sui legami a idrogeno e note come interazioni terziarie. Questi legami possono realizzarsi tra due o tre basi che generalmente non vengono considerate complementari, tra una base e lo scheletro formato dai gruppi fosfato e dallo zucchero ribosio della catena dell'RNA di trasporto o addirittura tra differenti parti dello stesso scheletro. È di particolare interesse il fatto che parecchie interazioni terziarie nel t-RNA riguardino i gruppi ossidrilici (OH) dello zucchero ribosio, dato che questi gruppi sono assenti dalle molecole di zucchero del DNA. Ciò vuol dire che queste interazioni non sono necessarie in una catena nucleotidica lineare regolare, quale è quella del DNA, mentre sono essenziali per stabilizzare l'avvolgimento della catena polinucleotidica del t-RNA.

Una disposizione insolita dei legami a idrogeno è stata riscontrata nello stelo o braccio accettore, in cui la coppia di basi nucleotidiche guanina-uracile compare al posto della coppia normale guanina-citosina o adenina-uracile. La possibilità di un simile appaiamento era stata suggerita parecchi anni prima, quando Francis H.C. Crick aveva osservato che altri tipi di appaiamento delle basi si sarebbero potuti probabilmente trovare in corrispondenza della terza base nell'interazione tra codone dell'RNA messaggero e anticodone dell'RNA di trasporto. Una delle disposizioni (non convenzionali) che Crick aveva postulato era proprio la coppia guanina-uracile, che sarebbe stata legata da due legami a idrogeno come risultato di una «oscillazione», o lieve spostamento laterale, di una delle due basi. Successive analisi e perfezionamenti a carico della mappa di densità elettronica, con una risoluzione di 0,25 nanometri hanno confermato questa oscillazione nell'appaiamento guanina-uracile in corrispondenza del braccio accettore.

Parecchi altri nuovi tipi di disposizione dei legami a idrogeno sono stati scoperti nelle interazioni terziarie tra base e base nella molecola di RNA di trasporto (*si veda la figura nella pagina a fronte*).

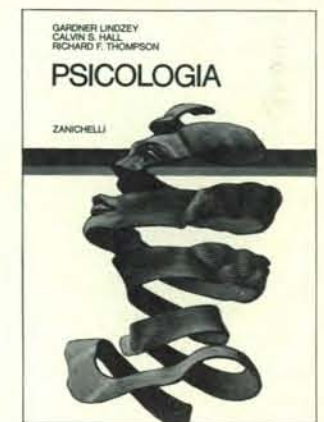
La maggior parte delle basi nucleotidiche appiattite presenti nell'RNA di trasporto sono disposte in due pile che formano i bracci della molecola a forma di *L*. Questa disposizione spiega l'insolita stabilità del t-RNA. Se si scalda una soluzione contenente molecole di questo acido nucleico, esse si denaturano, cioè la catena polinucleotidica che le costituisce si svolge e assume nella soluzione conformazioni casuali. Non appena si riraffredda, la molecola ritorna però rapidamente alla sua conformazione originaria. Questo comportamento è diverso da quello mostrato dalla maggior parte delle proteine, che si denaturano in maniera irreversibile: l'ovalbumina, per esempio, diventa bianca e opaca quando l'uovo viene bollito e rimane tale anche quando si raffredda.

Perché la molecola dell'RNA di trasporto ritorna così prontamente alla sua struttura originaria? È noto che l'interazione tra le due pile di basi nucleotidiche adiacenti, all'interno della doppia elica del DNA, è uno dei principali fattori di stabilizzazione di questa molecola. Analogamente, le basi del t-RNA sono prevalentemente idrofobe (cioè idrorepellenti) per cui si ritraggono dal solvente circostante, portandosi all'interno della catena polinucleotidica ripiegata: questo comportamento favorisce il ritorno della molecola di t-RNA alla sua conformazione originaria, cioè più stabile. Nelle proteine non esiste normalmente una interazione paragonabile a questa, per far ripiegare spontaneamente la catena polipeptidica. Risulta così che la struttura del t-RNA si organizza in questo modo per poter preservare l'aspetto stabilizzatore delle interazioni tra le basi impilate. Contemporaneamente, un certa architettura molecolare molto complessa mantiene le due pile di basi disposte ad angolo retto l'una rispetto all'altra.

Un aspetto importante delle interazioni terziarie che si riscontrano nel t-RNA per la fenilalanina ottenuta da lievito è il fatto che molte di esse interessano basi che sono sempre le stesse nelle sequenze polinucleotidiche di tutti i t-RNA. Inoltre, le basi che sono presenti in quelle regioni della catena polinucleotidica che possiedono un numero variabile di nucleotidi generalmente non sono sovrapposte, ma sono situate in anse che sporgono dalla superficie della molecola. Questi dati fanno pensare che l'impalcatura strutturale del t-RNA per la fenilalanina ottenuta da lievito possa ospitare le sequenze nucleotidiche che si trovano in altri t-RNA. Per esempio, nel fenilalanil-t-RNA di lievito una regione variabile dell'ansa *D* contiene due nucleotidi e questo segmento della catena polinucleotidica si inarca e poi si ripiega. Se vi fossero più nucleotidi in esso, probabilmente la convessità sarebbe più voluminosa; per contro, se i nucleotidi fossero meno, essa sarebbe più piccola. La dimensione di queste anse variabili non influenzerebbe, tuttavia, la ripiegatura complessiva della molecola.

Un certo numero di problemi importanti, riguardanti la struttura tridimensionale degli RNA di trasporto in generale, resta ancora insoluto. Non è chiaro, per esempio, quale sia la struttura dettagliata per i t-RNA con anse variabili molto grosse. È pure interessante la struttura dei t-RNA «iniziatori», cioè di quei t-RNA che danno inizio alla sintesi proteica, ponendo il primo amminoacido. Alcuni di questi t-RNA iniziatori hanno sequenze polinucleotidiche che si distaccano in una certa misura da quelle comuni agli altri t-RNA, in particolare per quanto riguarda l'ansa *T*. È probabile che differenze di questo tipo siano associate con una struttura lievemente diversa da quella dell'RNA di trasporto per la fenilalanina ottenuta da lievito.

I cristalli di questo t-RNA di lievito contengono quasi il 75 per cento di acqua ed è importante chiedersi se la molecola abbia la stessa forma in soluzione (dove è biologicamente attiva) e nei cristalli. Per fortuna sono state compiute numerose indagini sul t-RNA per la fenilalanina di lievito in soluzione ed esse hanno permesso di correlare la struttura osservata nei cristalli con varie caratteristiche chimiche della molecola. Per esempio, uno degli aspetti tipici della molecola in soluzione è che alcuni nucleotidi sembrano essere prontamente disponibili per una modificazione chimica quando alcuni reagenti vengono aggiunti alla soluzione, mentre altri nucleotidi non lo sono. Questa disparità sembrava inspie-

gabile fino a quando non si conobbe la struttura della molecola nel cristallo. A questo punto, è risultato chiaro che soltanto certi nucleotidi, quelli che sporgono dalla molecola allo stato cristallino, sono in grado di subire modificazioni chimiche. In generale esiste un'eccellente correlazione tra la suscettibilità di una regione della molecola di t-RNA alle modificazioni chimiche e l'accessibilità a quella regione della molecola, allo stato cristallino.

Parecchi altri tipi di esperimenti realizzati in soluzione possono venir interpretati alla luce della struttura tridimensionale. Sono inclusi in essi esperimenti basati sulla risonanza magnetica nucleare, che è sensibile alla struttura tridimensionale di una molecola. Numerosi ricercatori hanno trovato una valida correlazione tra i segnali di risonanza magnetica nucleare, che si ottengono da molecole di RNA di trasporto, e la struttura tridimensionale dedotta dall'analisi diffrattometrica con i raggi X del t-RNA legato alla fenilalanina di lievito allo stato cristallino. Questi e altri risultati confermano che la struttura della molecola di t-RNA allo stato cristallino è proprio la struttura della forma biologicamente attiva della molecola.

Le correlazioni di questo tipo sono importanti perché la ragione principale che spinge a determinare la struttura di una molecola sta nel riuscire a capire come essa funzioni in un sistema biologico. Che cosa ci ha insegnato la struttura del t-RNA sul modo in cui opera questa molecola? Su questo punto le risposte non possono essere precise perché non sono stati ancora realizzati gli esperimenti necessari. Innanzitutto si vorrebbe sapere come l'enzima amminoacil-t-RNA-sintetasi riconosce e seleziona soltanto il giusto t-RNA per l'attacco di un amminoacido specifico. Se si vuole capire a fondo questo processo, bisognerà determinare la struttura tridimensionale della sintetasi, quando forma un complesso con il t-RNA, in modo da poter osservare le interazioni specifiche tra l'enzima e l'acido nucleico. Studi di questo genere si stanno compiendo in vari laboratori e le risposte si dovrebbero avere tra non molto.

Un altro interrogativo importante è il seguente: perché la molecola di t-RNA ha una forma a *L*, in cui l'anticodone dista più di 7,6 nanometri dall'amminoacido attaccato? La risposta definitiva non è ancora stata data, ma è probabile che la forma di questa molecola sia correlata con la sua funzione essenziale di trasporto all'interno del ribosoma. Può darsi che la cellula abbia dovuto mettere a punto una molecola di t-RNA piegata in quel modo particolare, perché i due t-RNA adiacenti nei siti *A* e *P* potessero essere ravvicinati sul filamento di RNA messaggero, cosicché la catena polipeptidica in formazione fosse trasferita dall'uno all'altro. Forse il braccio accettore di *L* ruota nel ribosoma in modo che la catena proteica possa venir trasferita al t-RNA legato al codone successivo dell'RNA messaggero. Questo secondo t-RNA attaccherebbe la catena al proprio amminoacido e, a sua volta, la passerebbe a un terzo. Anche la considerevole distanza tra la fine del braccio accettore e l'ansa dell'anticodone, nella molecola di t-RNA, può essere funzionalmente importante in quanto differenti proteine ribosomiche possono interagire simultaneamente con regioni del t-RNA per cercare di conservare la precisione della sintesi proteica.

La visione che si è ottenuta della molecola di t-RNA, partendo da analisi diffrattometriche con i raggi X, compiute su cristalli, è essenzialmente statica. Nel suo ambiente naturale, all'interno della cellula, la molecola può subire modificazioni conformazionali, particolarmente quando interagisce con grosse strutture molecolari quali i ribosomi. Recenti esperimenti suggeriscono che, all'interno dei ribosomi, l'ansa *D* e l'ansa *T* della molecola di t-RNA si allontanino l'una dall'altra quando la molecola si sposta dal sito *A* al sito *P*. Può anche darsi che la forma dell'ansa anticodone si modifichi quando essa entra in contatto con l'RNA messaggero, entro il ribosoma.

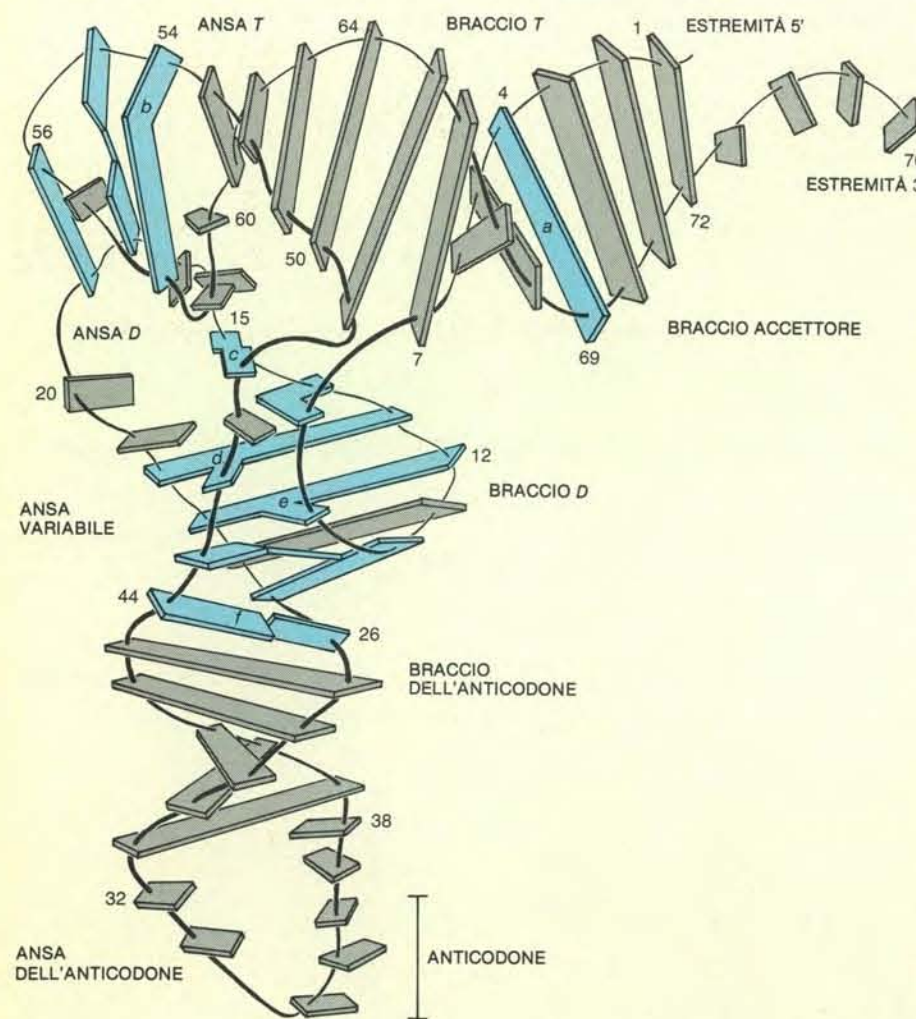
All'inizio di questo articolo abbiamo descritto il ruolo dell'RNA di trasporto nella sintesi proteica e lo abbiamo fatto con un certo dettaglio perché si tratta della funzione più essenziale di quella molecola nei sistemi biologici. Senza il t-RNA, l'informazione genetica non troverebbe espressione nella sintesi proteica. Inoltre esso partecipa a una varietà di altri processi, alcuni dei quali di notevole importanza. Per esempio, le sue molecole cedono amminoacidi a molecole proteiche preformate o alla struttura molecolare della parete cellulare dei batteri, indipendentemente dai ribosomi.

Un altro processo a cui il t-RNA partecipa è il controllo dell'espressione genica. È noto che certi t-RNA con un amminoacido attaccato determinano, regolando la trascrizione dell'RNA messaggero, se i geni, cioè i segmenti di DNA, potranno esprimersi o no. Il meccanismo particolareggiato non è noto, ma in alcuni sistemi questa funzione di controllo si associa con un nucleotide modificato particolare, presente nella molecola di t-RNA, per esempio un uracile trasformato in pseudouracile. Si pensa che il t-RNA favorisca il controllo dell'espressione di molti geni diversi, anche se di questi non si conosce il numero esatto. I casi che sono stati meglio studiati sono quelli dei geni che regolano la sintesi degli amminoacidi, in cui il t-RNA svolge un ruolo importante.

Altre osservazioni suggeriscono che l'RNA di trasporto possa essere implicato in ulteriori tipi di regolazione biochimica. Per esempio, nel corso dello sviluppo embrionale, un tipo di modificazione di certi nucleotidi in un t-RNA dà luogo a un altro tipo. Analogamente, quando una cellula normale diventa cancerosa, le modificazioni nucleotidiche a carico dei suoi t-RNA si alterano sostanzialmente. Non si sa ancora se queste trasformazioni sono associate o no alle funzioni regolatrici del t-RNA.

Un'altra questione tuttora aperta riguarda il numero relativamente elevato di modificazioni che si verificano nelle sequenze nucleotidiche dell'ansa *D* e in quelle dell'ansa variabile. Perché la natura ha fatto in modo di far variare i nucleotidi che sporgono dalla superficie della molecola di t-RNA? Si pensa in genere che queste sequenze non siano necessarie per la specificità della sintesi proteica: al contrario, potrebbero essere interessate nelle funzioni regolatrici delle molecole di t-RNA, dato che le regioni variabili potrebbero fornire i siti per il riconoscimento specifico da parte di altre molecole.

Infine, il t-RNA è associato non solo alla sintesi delle catene polipeptidiche, ma anche con a quella delle catene polinucleotidiche; quest'ultima avviene a opera di speciali enzimi come la transcriptasi inversa, che è stata scoperta alcuni anni fa come costituente di parecchi virus tumorali. Essa catalizza la sintesi di un filamento di DNA a partire da uno stampo costituito da un filamento singolo di RNA, con un flusso d'informazioni che procede in senso inverso rispetto al normale. Cosa sorprendente, un tipo specifico di t-RNA si lega dapprima all'enzima e all'RNA virale e segnala che la sintesi della coppia del DNA può iniziare. Non si sa assolutamente perché un t-RNA debba servire a questo scopo.

È probabile che molecole simili ai t-RNA fossero una componente essenziale dei primissimi organismi viventi. Una volta formate, la loro insolita instabilità può aver fatto sì che fossero gradatamente utilizzate, per scopi diversi dalla loro principale funzione, nella sintesi proteica. La conoscenza della struttura tridimensionale del t-RNA ha rappresentato un'importante progresso; tuttavia, rimangono ancora molte questioni da risolvere a proposito di questa versatile molecola e dei suoi molteplici compiti entro la cellula.

**Rappresentazione schematica dell'orientamento delle basi nella molecola di RNA di trasporto. Lo scheletro polinucleotidico è ridotto a una linea sottile, con le strutture simili a corte assicelle che stanno a indicare le basi non appaiate, mentre quelle più lunghe indicano coppie di basi. Le lettere si riferiscono alle strutture molecolari di pagina 48. Si noti la presenza di interazioni terziarie tra le tre basi della porzione centrale della molecola, sotto l'angolo della *L*. La molecola è nel complesso costituita da due colonne di basi disposte ad angolo retto l'una rispetto all'altra. Le interazioni tra basi parallele costituiscono un importante forza stabilizzatrice.**

# Una nuova teoria della visione dei colori

*Oltre 25 anni di esperimenti sembrano indicare che l'identificazione dei colori si basa su triplette di informazioni relative alla riflettanza degli oggetti che vengono elaborate da un sistema retina-corteccia*

di Edwin H. Land

La tradizione scientifica di semplificare le condizioni in cui vengono effettuati gli esperimenti ci ha lasciato fino a poco tempo addietro senza una spiegazione soddisfacente di come l'occhio vede i colori nella vita di tutti i giorni. Paradossalmente la moderna tecnologia della fotografia a colori ha rafforzato la convinzione che i colori osservati da Newton nello spettro sono, con varianti di poca importanza, i colori del mondo che ci circonda. Noi sappiamo, per esempio, che se usiamo una pellicola a colori per luce diurna per fotografare una scena illuminata da una lampada a filamento di tungsteno, l'immagine sarà caratterizzata da una forte dominanza rossastra. Questo avviene, diciamo, perché la luce della lampada è troppo «rossa», senza chiederci come mai ci è possibile cambiare l'illuminazione della scena senza osservare alcun cambiamento di colore degli oggetti familiari: mele, limoni, fragole, pane, facce umane (oggetti i cui colori sono così difficili da riprodurre correttamente su uno schermo televisivo a colori).

Come fa allora l'occhio a non tener conto dell'eccesso di «rosso» in una stanza illuminata con lampade comuni? Come spero di dimostrare in questo articolo, nel determinare i colori l'occhio non percepisce affatto l'eccesso di rosso perché tale determinazione non dipende dal flusso di energia raggiante che giunge all'occhio. L'occhio si è evoluto per vedere il mondo con colori stabili, indipendenti dall'illuminazione, che è sempre imprevedibilmente mutevole. Come sia possibile raggiungere questo risultato è una cosa che mi ha sempre affascinato.

Nel 1959 ho descritto una serie di esperimenti in cui un'immagine prodotta dalla proiezione in sovrapposizione di due diapositive in bianco e nero, proiettate l'una attraverso un filtro rosso e l'altra in luce bianca, viene percepita con quasi tutta la gamma di colori presente nell'originale (si veda *Experiments in Color Vision* di E.H. Land, in «Scientific American», maggio 1959). Per produrre una tale immagine, la diapositiva proiettata attraverso il filtro rosso viene ripresa attraverso lo stesso filtro e la diapositiva proiettata in luce bianca viene ripresa attraverso un filtro verde. Ci si attenderebbe che la proiezione delle due immagini sovrapposte sia solo in bianco, rosso e vari toni di rosa. E invece si osserva un'immagine notevolmente simile a quella a pieni colori riprodotta nella pagina a fronte. Nella proiezione in luce bianca e rossa i peperoni sono verdi, i ravanelli e le fragole sono rossi, l'arancia è arancione, le banane e il limone sono giallo chiaro, la tavoletta di legno e il manico del coltello sono marrone e il disegno sul piatto è blu.

A partire da questi vecchi esperimenti siamo risaliti passo per passo, nel corso di vent'anni, alla spiegazione di come il nostro sistema visivo riesca a estrarre informazioni consistenti sui colori del mondo che ci circonda, un mondo in cui praticamente ogni scena è illuminata in maniera non uniforme, dove la composizione spettrale della luce può variare in maniera enorme e dove illuminazioni brevi come un lampo sono sufficienti per un'identificazione accurata dei colori. Se i fotorecettori della retina funzionassero anche solo approssimativamente come ci hanno insegnato a scuola, e cioè come rivelatori d'intensità della radiazione con sensibilità centrata in tre diverse zone dello spettro, noi confonderemmo continuamente un colore con l'altro. Un oggetto che sembrava giallo in una parte del nostro campo visivo potrebbe apparire verde o grigio o anche rosso qualora venisse spostato in un'altra parte. E invece gli oggetti conservano il loro colore in una gran varietà di condizioni di luce. Questa costanza non è un effetto secondario, è invece talmente fondamentale da condurre a una nuova descrizione del meccanismo della visione dei colori.

I pigmenti della visione sono molecole fotosensibili che rispondono a una banda molto ampia di lunghezze d'onda. I tre pigmenti che si trovano nei coni della retina coprono l'intero spettro visibile con tre bande d'assorbimento ampie e ampiamente sovrapposte. Il pigmento con massimo di sensibilità alla lunghezza d'onda di 440 nanometri risponde in un certo grado a tutta la metà dello spettro di più bassa frequenza. Gli altri due pigmenti rispondono a circa due terzi dello spettro con massimi di sensibilità separati soltanto di 30 nm, essendo situati a 535 e 565 nm (*si veda la figura in alto a pagina 54*).

Nel discorso che segue i nomi dei colori - rosso, blu, verde e così via - verranno utilizzati per descrivere la sensazione dei colori che abbiamo guardando il mondo che ci circonda. In breve, solo i nostri occhi possono classificare il colore degli oggetti, gli spettrofotometri non possono farlo. Questo punto è importante perché molte persone che assistono per la prima volta a qualcuno dei nostri esperimenti identificano un colore come rosso o verde ma poi chiedono subito di che colore si tratta in realtà, come se i loro occhi venissero ingannati. La risposta è che l'occhio non viene ingannato. Sta funzionando esattamente come deve, con involontaria precisione, vedendo colori costanti in un mondo illuminato da flussi di radiazione mutevoli e imprevedibili.

Dato che sono convinto che lo studio del colore nelle immagini colorate si cominci più fruttuosamente esaminando immagini del tutto prive di colore e non complicate da questa presenza, descriverò in dettaglio un'esperienza interessante in questo senso. Il sistema fotorecettore ipersensibile basato sui bastoncelli della retina funziona con una illuminazione 1000 volte più debole del sistema basato sui coni. Ciò consente di rispondere a una domanda interessante: che colori si vedono se viene attivato solo il sistema a bastoncelli? Una procedura possibile consiste nell'uso di un paio di occhiali ben schermati e muniti di lenti grige che riducono 30 000 volte l'illuminazione incidente. Dopo una mezz'ora da che si indossano gli occhiali, diviene visibile anche l'interno di una stanza illuminata normalmente, al livello tipico di 200 lux. L'illuminazione efficace nella stanza corrisponde quindi, per chi indossa gli occhiali, a 1/150 di lux. Guardandosi attorno, i familiari oggetti colorati appaiono privi di colore, con una gamma di luminosità che va dal bianco al nero e con un aspetto simile a quello che assumono in una fotografia in bianco e nero ripresa attraverso un filtro verde per selezioni cromatiche. In altre parole i rossi appaiono molto scuri, i verdi più chiari, i blu scuri, i bianchi luminosi e i neri molto scuri.

In questo mondo privo di colori si scopre che la natura dell'immagine non è determinata dal flusso di energia raggiante che raggiunge l'occhio. L'illuminazione può infatti venire facilmente mo-

**La riproduzione di questa natura morta, utilizzata per riprendere le quattro immagini in bianco e nero della figura in basso, è stata ottenuta con metodi usuali di fotografia a colori e di fotoincisione, per mostrare al lettore i colori originali degli oggetti fotografati. Le riprese fotografiche in bianco e nero sono state ottenute con combinazioni di pellicole e di filtri tali da riprodurre fedelmente le diverse sensibilità spettrali dei quattro sistemi di fotorecettori presenti nella retina: i tre sistemi di cellule a cono e il sistema ipersensibile a bastoncelli.**

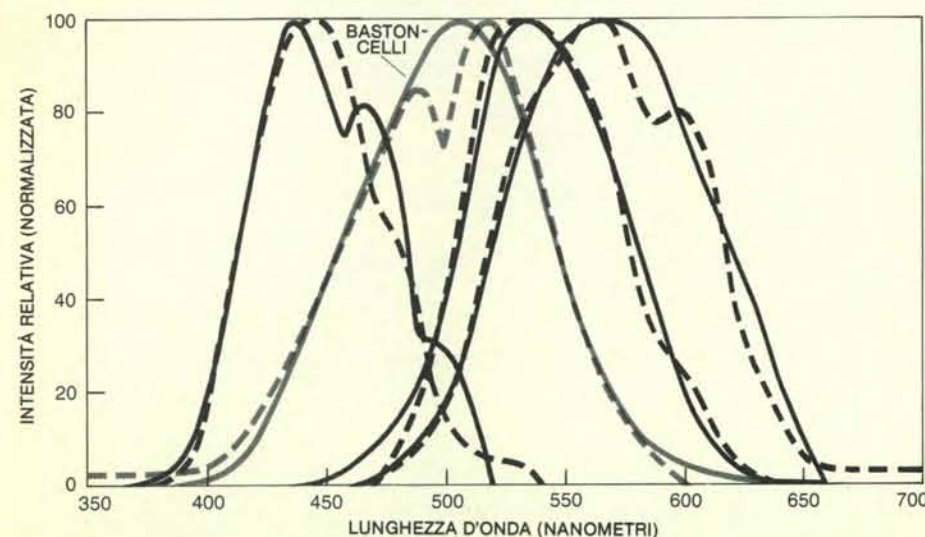
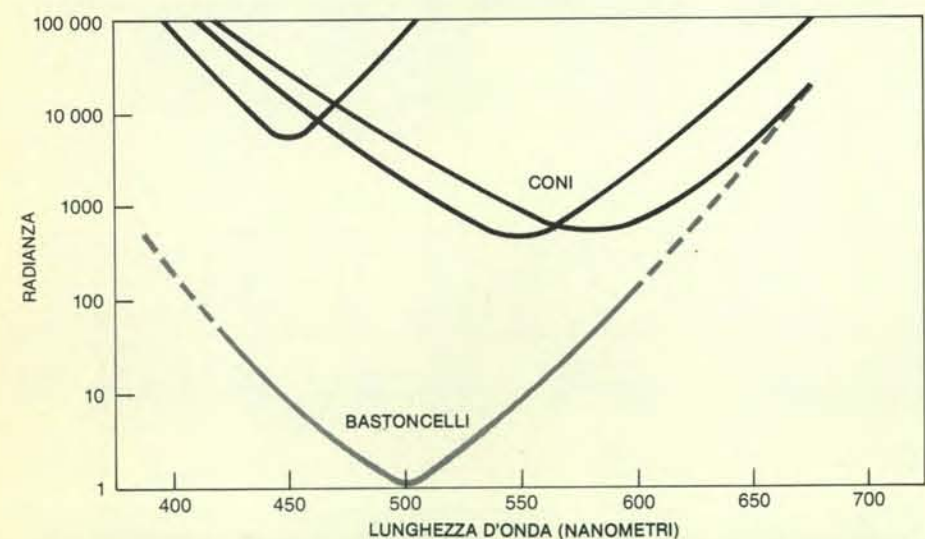
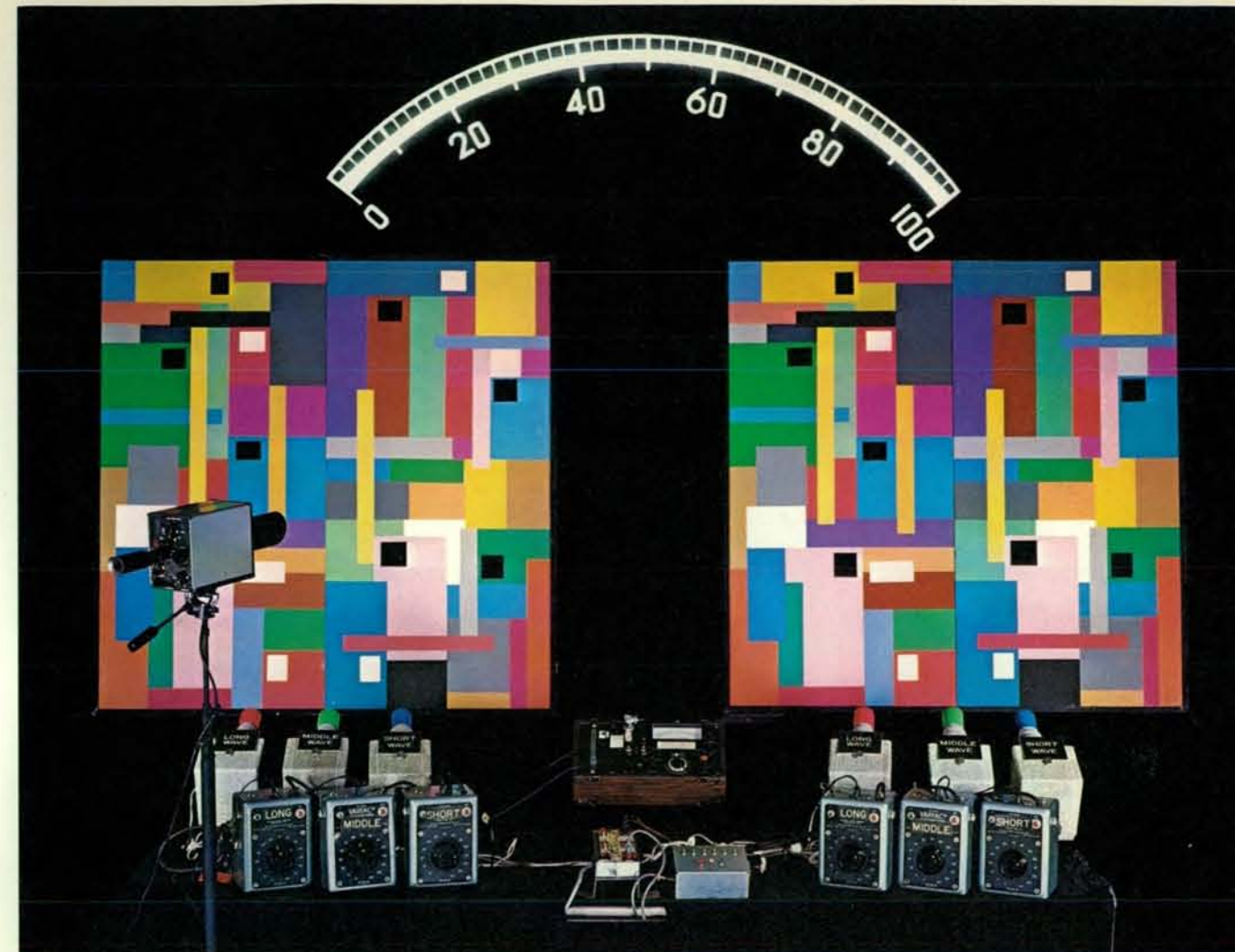
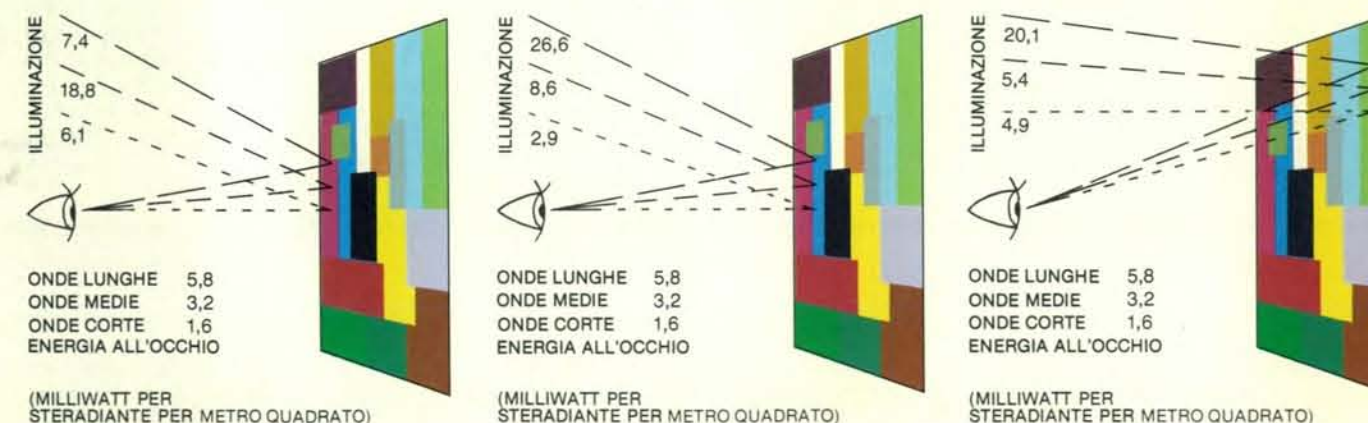
**Queste immagini in bianco e nero di una natura morta sono quelle che l'autore chiama fotografie retinex. L'immagine in alto a sinistra è stata presa con una pellicola la cui sensibilità spettrale è stata alterata in modo da renderla uguale a quella dei coni sensibili a grandi lunghezze d'onda. Questa fotografia consente a chi la guarda di vedere un'immagine priva di colori che approssima quella prodotta dal sistema di coni suddetto. L'immagine in alto a destra mostra come la stessa natura morta viene vista dai coni con pigmento sensibile a lunghezze d'onda intermedie. L'immagine in basso a sinistra corrisponde alla visione dei coni sensibili a piccole lunghezze d'onda e quella in basso a destra all'immagine vista dai bastoncelli. A differenza delle immagini prodotte dai coni, che non possono essere viste separatamente, le immagini prodotte dai pigmenti dei bastoncelli possono venire studiate isolatamente a bassissimi livelli di illuminazione; a tali livelli infatti non si hanno interferenze da parte dei sistemi di coni, caratterizzati da una sensibilità molto inferiore.**

dificata in modo che da una superficie che continua ad apparire molto scura vi sia un flusso maggiore di quello proveniente da una zona che rimane molto chiara. E questo indipendentemente dal fatto che queste regioni siano oggetti tridimensionali o artefatti nascosti in un montaggio di pezzi di carta chiara e scura. Di qui nasce il paradosso che gli oggetti, per esempio i pezzi di carta, indipendentemente dal fatto che siano chiari, scuri o di colore intermedio, mantengono il loro aspetto senza variazioni significative quando vengono spostati in zone in cui è diverso il flusso incidente. I fogli di carta chiari continueranno a venire percepiti chiari e quelli scuri come tali, anche se l'occhio riceve da essi il medesimo flusso di radiazione. Forti variazioni di flusso all'interno del campo visivo vengono percepiti poco o niente.

E ancora, in un complicato collage di carte di varie dimensioni, forme e potere riflettente, la luminosità di una data zona non pare variare sensibilmente quando viene spostata da una parte all'altra del collage e quindi associata arbitrariamente all'ambiente circostante. Quando una piccola superficie viene circondata da una molto più grande, allora la percezione della luminosità della prima cambia leggermente in funzione del fatto che la superficie più grande sia più chiara o più scura. In generale, comunque, il fatto importante è che la percezione della luminosità di una certa superficie non viene modificata dall'ambiente immediatamente circostante e neppure dalle superfici più grandi che la circondano.

Benché finora sia stato descritto il comportamento dei soli fotorecettori ipersensibili a bastoncello, tutte le osservazioni fatte sulla stabilità della percezione della luminosità possono venire ripetute con un collage di pezzi di carta bianchi, grigi e neri osservato a livelli di illuminazione ordinari. Poniamo per esempio un pezzo di carta nera opaca o, meglio, un pezzo di velluto nero su un lato del collage e un pezzo di carta bianca sul lato opposto, a parecchia distanza dal primo, con in mezzo una serie di carte chiare e scure. Ora è possibile disporre una forte sorgente di luce tanto vicino al velluto nero che questo riflette verso l'occhio più luce di quanta non ne rifletta la carta bianca più lontana dalla luce e, malgrado questo, il velluto continuerà a sembrare nero e la carta bianca. In effetti, con il collage sempre illuminato lateralmente in maniera non uniforme, l'impressione di bianco e di nero viene mantenuta muovendo il velluto e la carta in tutte le posizioni possibili.

Questa notevole abilità dell'occhio nel distinguere il potere riflettente degli oggetti, indipendentemente dal flusso di radiazione che lo raggiunge, è la base su cui è possibile costruire una descrizione soddisfacente della visione dei colori. La prima risposta del sistema visivo alla luce incidente è l'assorbimento di questa da parte dei fotorecettori. Per quanto il segnale iniziale proveniente dai segmenti esterni delle cellule fotorecettrici sembri dipendere dal flusso di radiazione assorbito dal pigmento fotosensibile, la risposta finale complessiva del sistema visivo si dimostra praticamente indipendente dal flusso di radiazione assorbito dal pigmento e legata invece al potere riflettente dell'oggetto osservato.

Questo richiede un sistema di elaborazione dei flussi di radiazione, che può risiedere nella retina, nella corteccia cerebrale o in entrambe. Vista l'incertezza sulla localizzazione di tale sistema di elaborazione ho proposto di definire con il termine «retinex» l'insieme dei meccanismi biologici che consentono di convertire i flussi di radiazione incidenti sulla retina in informazioni relative al potere riflettente degli oggetti. Inoltre in questo articolo utilizzerò il termine «luminosità» per indicare la sensazione relativa alla brillanza delle superfici, sensazione che abbiamo visto dipendere più dal potere riflettente di queste che dalla loro effettiva brillanza, da cui dipende il flusso di radiazione che raggiunge la retina. Per quanto i bastoncelli possano venire eccitati con livelli di illuminazione al di sotto del valore di soglia per i coni, questi non possono venire eccitati senza stimolare anche i bastoncelli. La luminosità delle immagini prodotte dai vari tipi

**Le sensibilità spettrali normalizzate dei quattro pigmenti visivi (*linee continue*) coprono lo spettro visibile sovrapponendosi ampiamente. La curva con massimo a 550 nanometri corrisponde alla sensibilità del pigmento dei bastoncelli. Le altre curve corrispondono ai pigmenti dei coni. Le curve tratteggiate mostrano la sensibilità delle combinazioni filtro-pellicola scelte per simulare quelle dei pigmenti della retina e utilizzate per le fotografie retinex in bianco e nero. Le curve relative ai coni sono state adattate dai lavori di Paul Brown e George Wald della Harvard University. La curva relativa ai bastoncelli è la curva standard di sensibilità scotopica.**

**Le soglie di eccitazione dei recettori della retina sono molto diverse. Il sistema ipersensibile a bastoncelli consente la visione a livelli di illuminazione 1000 volte più deboli di quelli necessari per attivare i sistemi a coni. Nel laboratorio dell'autore è stato dimostrato che gli oggetti colorati sono visti nella quasi completezza dei loro colori, a livelli d'illuminazione aggiustati in modo da eccitare solo il sistema a bastoncelli e quello a coni sensibili a grandi lunghezze d'onda.**

**Nell'esperimento dei collage colorati sono utilizzati due collage identici di pezzi di carta, montati su tavole quadrate di 1,35 metri di lato. Le carte colorate sono opache per minimizzare le riflessioni speculari. Ogni collage viene illuminato da una terna di proiettori, muniti di filtri passa-banda e di controlli indipendenti di luminosità, in modo da poter mescolare nelle proporzioni volute i fasci di luce a grande, media e piccola lunghezza d'onda («rossa», «verde» e «blu»). Si può puntare un fotometro su qualsiasi zona del collage per misurare, singolarmente nelle tre bande, il flusso di radiazione che tale zona riflette verso l'occhio. La lettura del fotometro è proiettata sulla scala sopra i due collage. In un esperimento tipico i proiettori vengono aggiustati in modo che la zona bianca del collage di sinistra e una zona verde (o qualsiasi altra zona) nel collage di destra riflettano entrambe la medesima tripletta di energie nelle tre bande dello spettro. La situazione reale non può venire riprodotta qui, a causa delle limitazioni della tecnica di riproduzione dei colori. Nella realtà la zona bianca continua a essere bianca e quella verde continua a essere verde, malgrado l'occhio riceva la medesima tripletta di flussi da entrambe.**

**Negli esperimenti con i collage flussi energetici identici incidenti sulla retina provocano differenti sensazioni di colore. In questo esempio, con i proiettori di luce a grande, media e corta lunghezza d'onda regolati come indicato, una superficie rossa continua a venire vista rossa (*a sinistra*), una blu rimane blu (*al centro*) e una verde rimane verde (*a destra*), anche se tutte e tre riflettono la medesima tripletta di energie nei fasci di luce a grande, media e piccola lunghezza d'onda. La medesima tripletta può essere fatta riflettere da qualsiasi altro punto del collage: se si tratta di una zona bianca, essa rimane bianca, se la zona è grigia rimane grigia, se la zona è gialla rimane gialla e così via.**

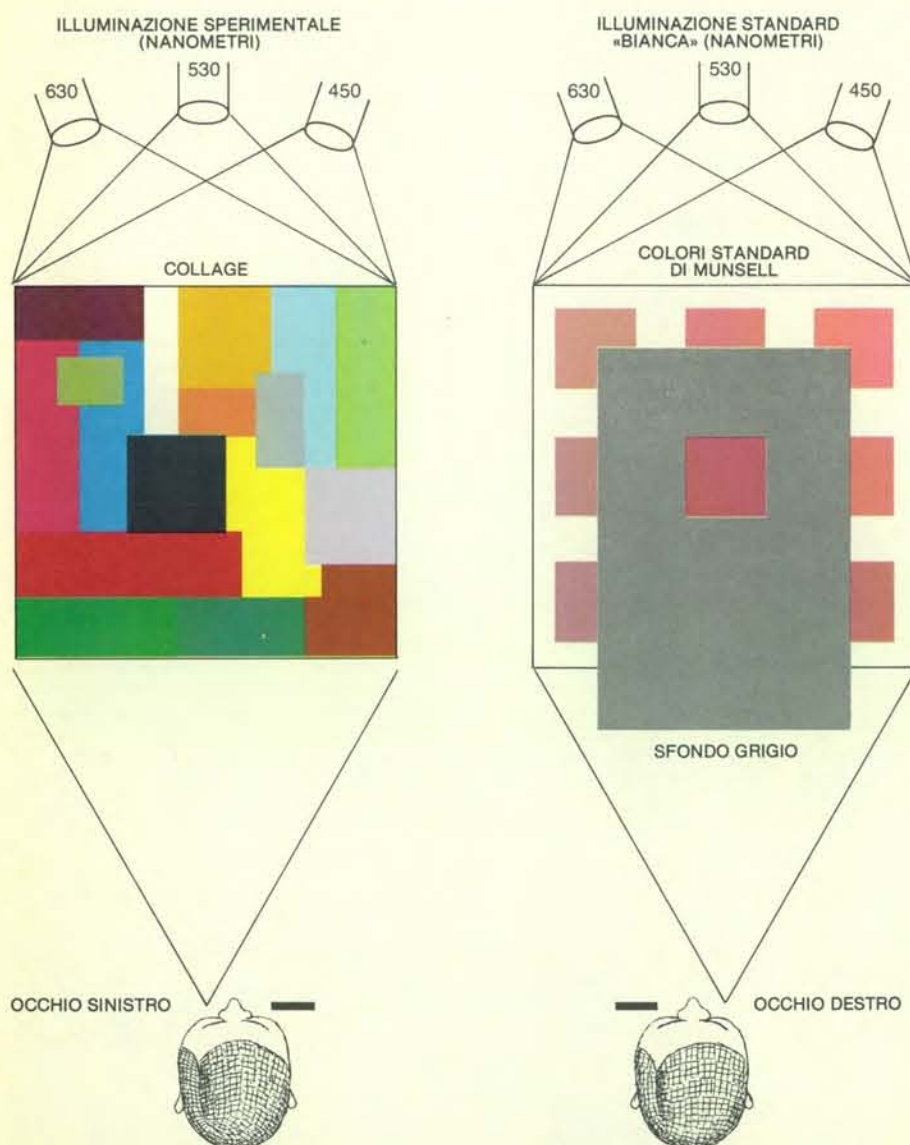
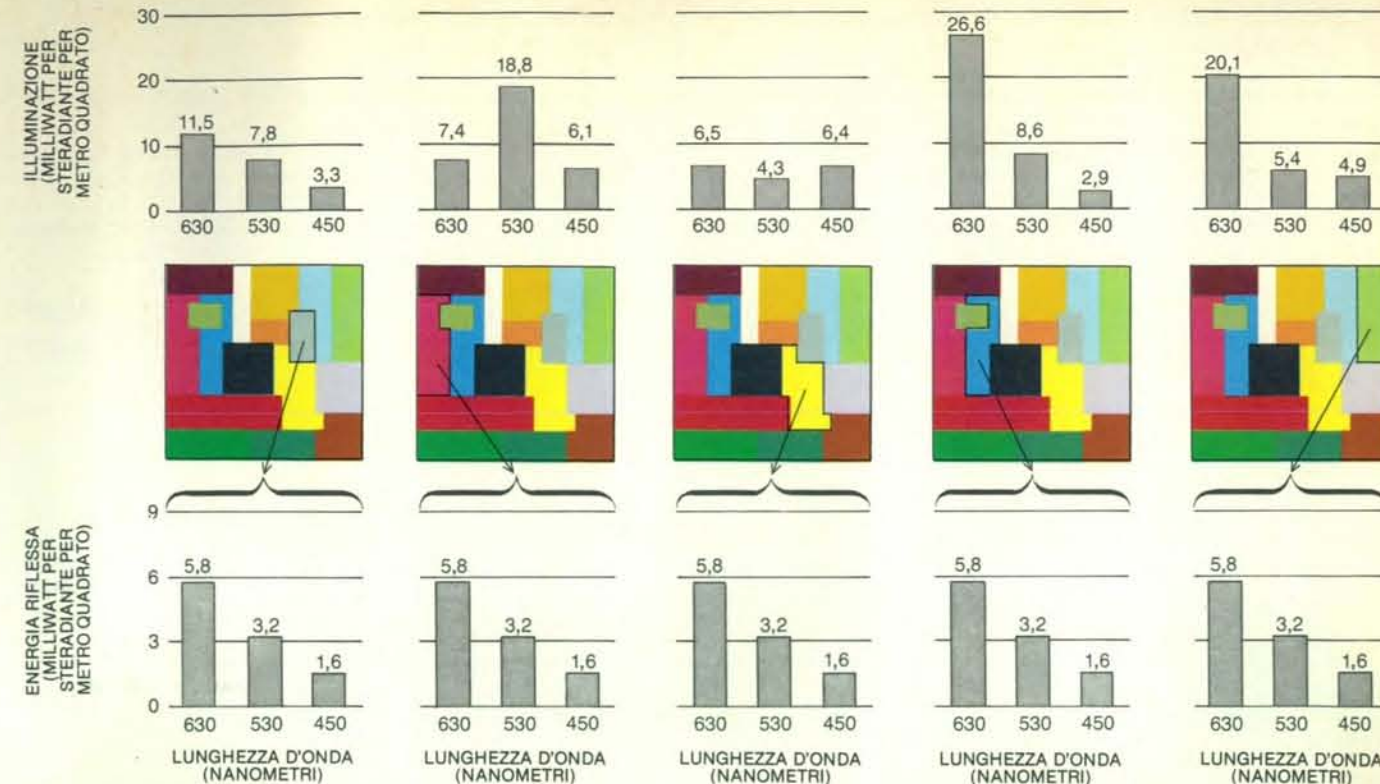
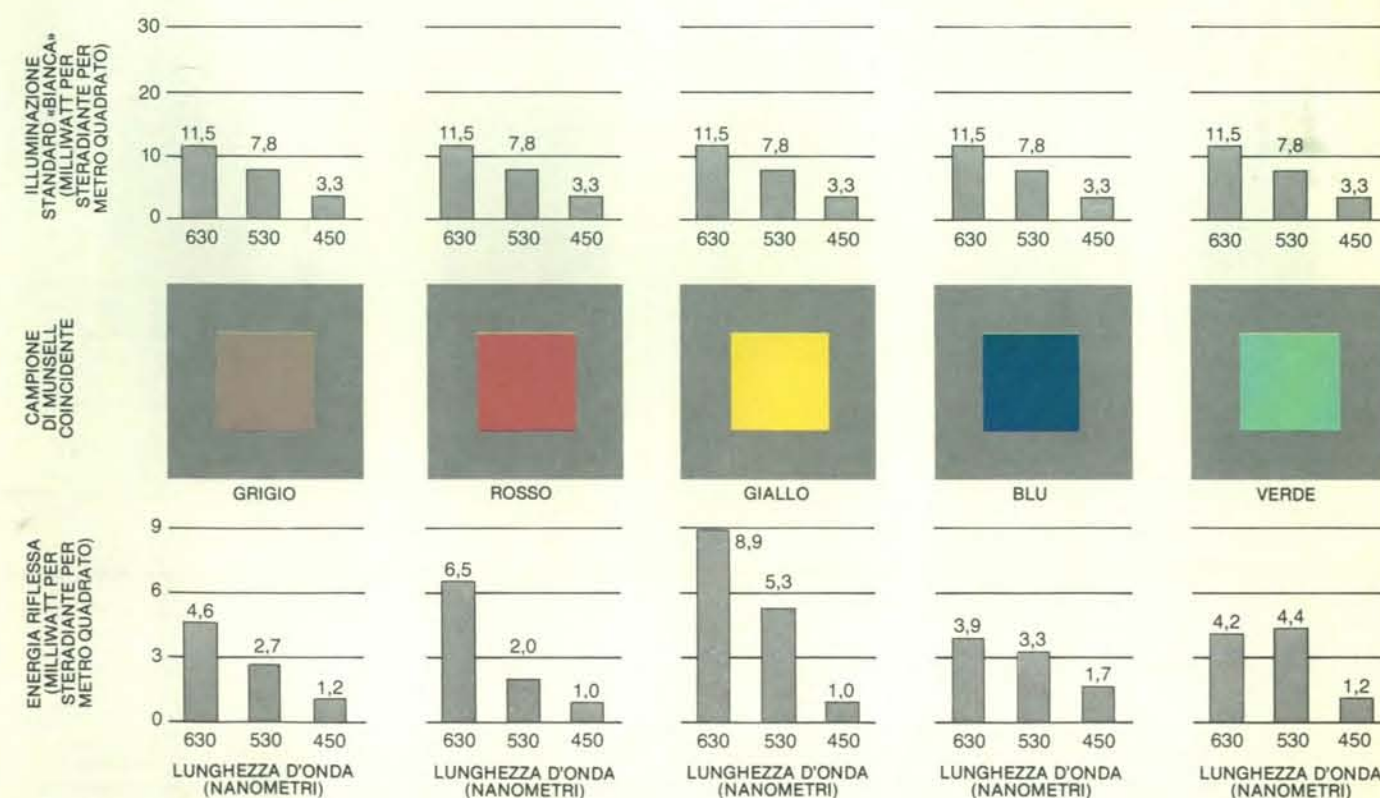
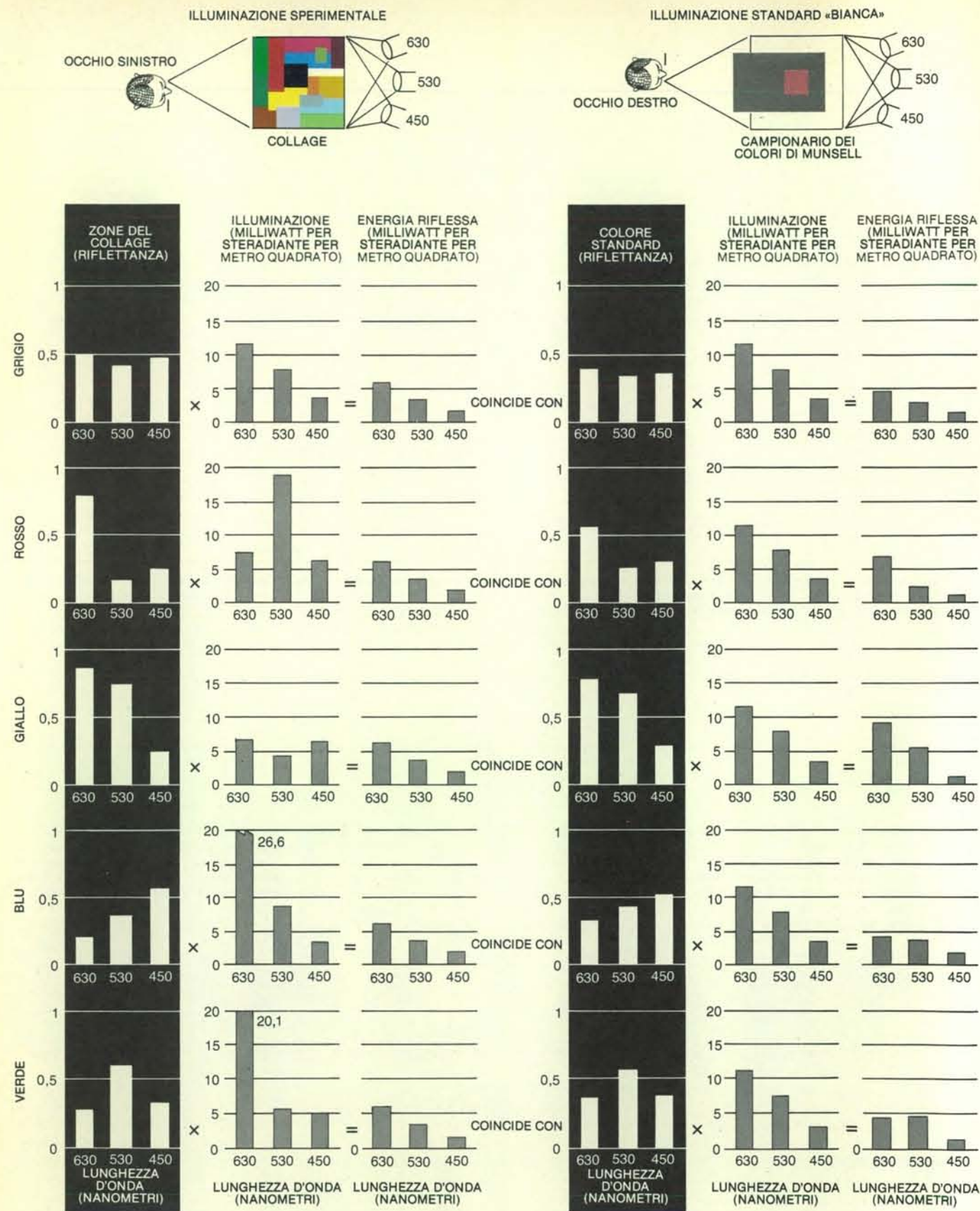
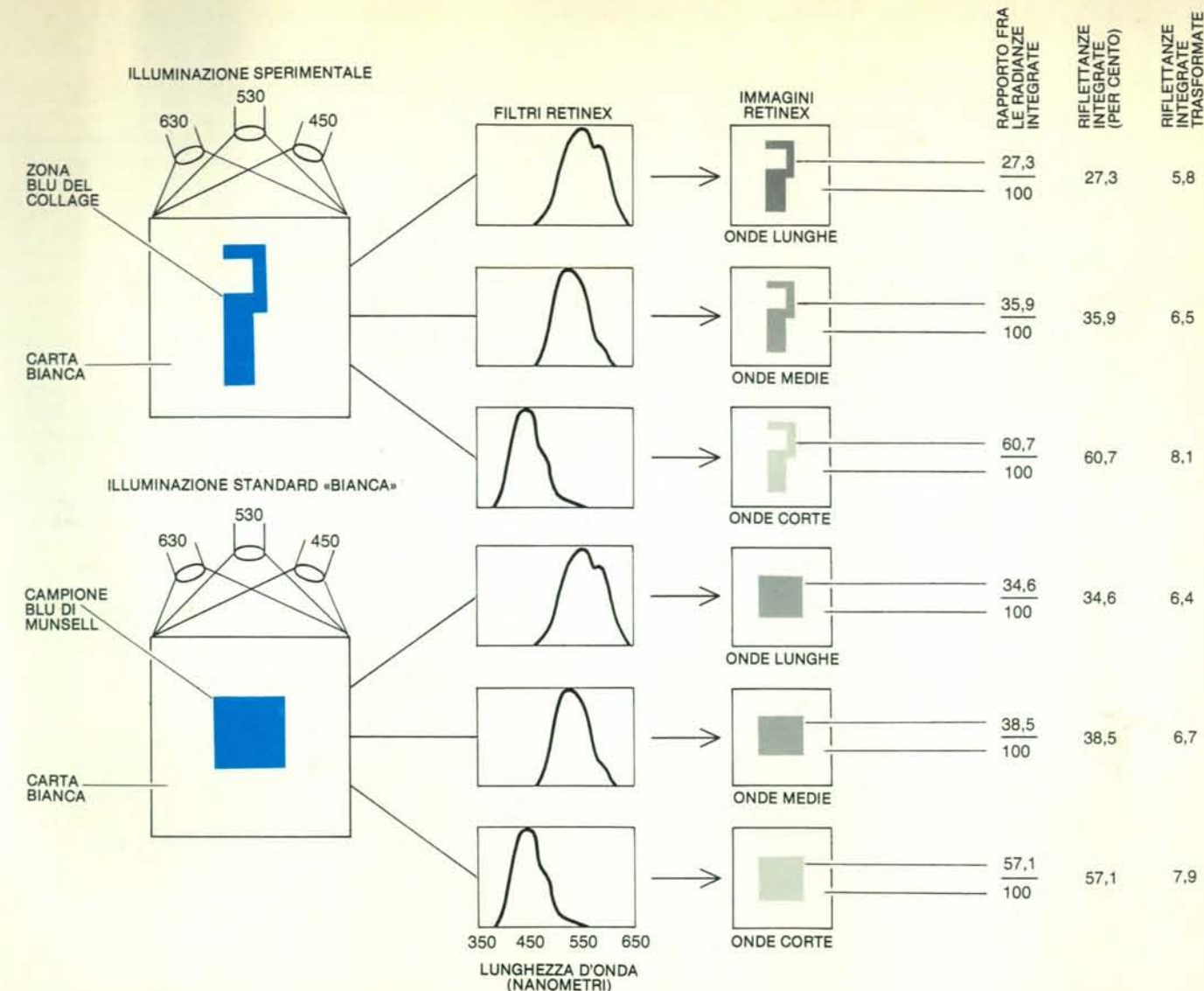
di coni può quindi venire studiata solo con il metodo fotografico che descriverò più avanti o mediante calcoli basati sui risultati di misure spettroradiometriche.

Ora che sappiamo come a bassi livelli di illuminazione un singolo sistema di recettori che funziona isolatamente fornisca un'immagine costituita solo da zone di diversa luminosità e priva di colore, sarebbe interessante vedere se è possibile mettere in azione insieme ai bastoncelli uno dei tre sistemi di coni, in modo da avere in funzione solo due sistemi di recettori. Questo esperimento è stato realizzato e fornisce una solida conferma delle idee suggerite da tutto il nostro lavoro precedente con stimoli binari in luce bianca e rossa e degli esperimenti successivi con stimoli ternari ottenuti mediante oggetti a più colori visti sotto un'illuminazione variabile. Questo esperimento, che sta diventando un classico, è stato ideato dai miei colleghi John J. McCann e Jeanne L. Benton.

L'esperimento si effettua illuminando un soggetto a più colori con una sorgente a banda stretta, centrata a 550 nanometri, di intensità appena superiore alla soglia necessaria per rendere il soggetto visibile all'occhio adattato al buio. In questo modo si è certi che viene eccitato solo il sistema ipersensibile a bastoncelli. Si aggiunge poi una seconda sorgente a banda stretta centrata a 656 nanometri e di intensità sufficiente per attivare il sistema a coni sensibile alle grandi lunghezze d'onda ma non quello sensibile alle lunghezze intermedie. In tali condizioni risultano eccitati solo due sistemi di fotorecettori, ossia i bastoncelli e i coni sensibili alle lunghezze d'onda maggiori.

L'immagine osservata è caratterizzata da un'ampia gamma di colori, tale da consentire all'osservatore di assegnare a ciascuna zona dell'oggetto il nome del colore che era possibile vedere con l'illuminazione normale. Il risultato ricorda quello ottenibile con la proiezione di immagini in luce bianca e rossa. Questo esperimento conferma la nostra vecchia ipotesi che le informazioni sulla luminosità raccolte a due diverse lunghezze di onda da due sistemi fotorecettori non vengono mediate punto per punto bensì sono mantenute distinte e confrontate. Sappiamo ora che il sistema a bastoncelli non consente da solo la visione dei colori e che neppure un'illuminazione a grande lunghezza d'onda consente da sola l'osservazione di un'immagine multicolore. Ma quando funzionano entrambi i sistemi si osserva una notevole varietà di colori, in particolare rossi, gialli, marroni, bluverdi, grigi e nero.

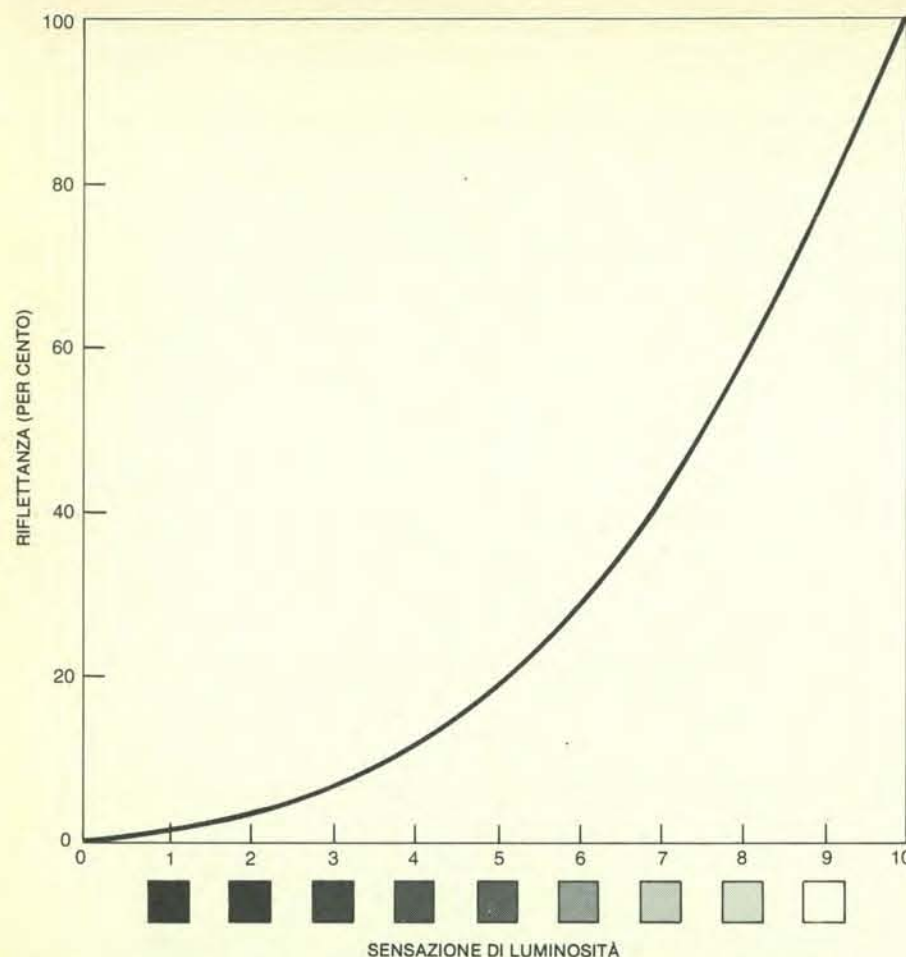
Allora che cosa produce la visione dei colori? La loro visione può venire ascritta a un processo operante in qualche punto del sistema visivo che confronta la luminosità di due immagini distinte in due diverse bande dello spettro fornite da due sistemi retinex indipendenti. L'esperimento con due sistemi fotorecettori rende possibile ipotizzare che, quando vengono associate tre immagini indipendenti che riportano le luminosità percepite dai tre sistemi di fotorecettori sensibili a piccole, medie e grandi lunghezze d'onda, la percezione dei colori sia dovuta al confronto delle diverse luminosità, zona per zona. La ragione per cui il colore di ogni punto dell'immagine risulta essenzialmente indipendente dal rapporto dei tre flussi di radiazione corrispondenti alle tre regioni spettrali di massima sensibilità dei recettori, va cercata nel fatto che il colore dipende solamente dalla luminosità in ognuna delle bande spettrali e la luminosità non dipende dal flusso.

Come abbiamo già visto, le sensibilità spettrali dei pigmenti visivi si sovrappongono ampiamente. Se illuminassimo una scena con l'intera banda di radiazioni cui è sensibile uno dei pigmenti, sarebbe possibile vedere una gran varietà di colori, in quanto verrebbe eccitato più di un sistema retinex. Con l'impiego di filtri e di emulsioni fotografiche appropriate, riesce peraltro possibile isolare le luminosità che altrimenti verrebbero integrate nella percezione del colore. Fotografie in bianco e nero riprese allo scopo si possono chiamare immagini retinex.

**Nell'esperimento di confronto dei colori si impiegano un collage semplificato con 17 zone di colore (*a sinistra*) e un campionario di colori standard, il *Munsell Book of Color*, che contiene 1150 campioni di colore (*a destra*). Il collage viene illuminato con tre sorgenti a banda stretta: una a 630 nanometri, una a 530 e una a 450. Le proporzioni dei tre fasci di luce si possono variare in modo che la tripletta di energie riflessa da una zona qualsiasi uguagli quella riflessa in precedenza da un'altra zona. In questo esperimento sono state scelte cinque zone, di colore grigio, rosso, giallo, blu e verde, dalle quali far riflettere in successione la medesima tripletta di energie. Nel corso di ognuna delle cinque parti consecutive dell'esperimento, l'osservatore doveva scegliere fra gli standard di colore quelli che più si avvicinavano al colore delle 17 zone del collage. Gli standard di colore, durante l'intero esperimento, erano illuminati con una mescolanza costante delle tre luci scelta in modo che il bianco del campionario di Munsell risultasse all'osservatore «il miglior bianco possibile». L'esperimento era condotto in modo che gli osservatori potessero vedere il collage con un occhio e con l'altro i colori standard circondati da una maschera grigia.**

**Le regolazioni dei proiettori a banda spettrale stretta usati per illuminare il collage semplificato negli esperimenti di confronto con il campionario dei colori di Munsell, indicate dalle barre in alto, sono state scelte in modo che cinque diverse zone del collage (*segnalate dalle frecce*) riflettessero, nelle cinque fasi successive dell'esperimento, le medesime triplette di energie: 5,8 unità di flusso per la radiazione di grande lunghezza d'onda, 3,2 per quella di media lunghezza d'onda e 1,6 per quella di piccola lunghezza d'onda. La figura in basso mostra i colori standard scelti sotto l'illuminazione costante come identici alle cinque zone del collage che avevano riflesso la stessa tripletta di energie.**

**Sono riprodotti gli standard di Munsell scelti come identici alle zone del collage che avevano riflesso le medesime triplette di energia. Gli standard erano illuminati con una miscela costante di strette bande spettrali (*barre in alto*) e osservati su uno sfondo grigio pure costante. L'energia riflessa dagli standard è indicata dalle barre in basso. La sensazione di identità degli standard e dei colori del collage non è dovuta all'identità del flusso di energia riflessa ma ai processi descritti nell'illustrazione successiva.**

**L'analisi degli esperimenti di confronto dei colori fornisce i primi indizi su che cosa consenta l'identificazione dei colori indipendentemente dai flussi di energia riflessi dalle zone del collage e dai corrispondenti colori standard. L'efficacia con cui una data zona del collage riflette la luce in ognuna delle tre bande spettrali (*prima colonna*) moltiplicata per l'energia incidente nelle rispettive bande (*seconda colonna*) fornisce la tripletta di energie che giunge all'occhio (*terza colonna*). Le tre colonne a destra contengono gli stessi dati per i colori standard scelti dagli osservatori come identici alle zone del collage. Mentre l'illustrazione in basso nella pagina precedente mostra come l'occhio non confronti i colori in base ai valori assoluti delle triplette di energia che giungono all'occhio, questa figura mostra che l'identificazione della corrispondenza avviene quando corrispondono le riflettanze delle due superfici, indicate nella prima e nella quarta colonna.**



**Il ruolo sostenuto dalla riflettanza e dalla corrispondente grandezza psicofisica, la luminosità, nel corso degli esperimenti di confronto dei colori è stato studiato mediante combinazioni di fotomoltiplicatori e filtri retinex aventi le medesime sensibilità spettrali dei pigmenti dei coni. Per ciascuna combinazione di fasci di luce (*in alto*) si è confrontata la radianza integrata in ognuna delle bande spettrali dei filtri con la corrispondente radianza integrata di un foglio bianco. Il rapporto fra le radianze integrate fornisce la riflettanza integrata di ognuna delle zone del collage, qui espressa in forma percentuale. Per il colore standard ritenuto identico si è determinata in maniera analoga una corrispondente serie di rapporti (*in basso*). Il passaggio finale nella determinazione di un equivalente fisico della luminosità è la determinazione della relazione fra le differenze di riflettanza integrata e di luminosità. Questa trasformazione è illustrata nella pagina seguente.**

Questa tecnica fotografica, che fa uso di emulsioni d'argento, ha due funzioni. Innanzitutto provvede sensibilità spettrali identiche a quelle dei pigmenti dei fotorecettori; inoltre fornisce immagini da far esaminare a osservatori umani. È il sistema visivo umano che traduce questa immagine fotografica in un insieme di luminosità. Sarebbe interessante poter esaminare l'immagine in bianco e nero con un solo sistema di coni, osservando quindi le luminosità tipiche per quel sistema. Invece in ogni punto dell'immagine la riflessione è praticamente la stessa per tutto lo spettro visibile. Ne segue che con una fotografia in bianco e nero noi stimoliamo tutti i recettori con la medesima informazione, cioè con le energie che verrebbero assorbite da uno solo dei pigmenti. Se ipotizziamo che tutti i sistemi retinex elaborino le informazioni nel medesimo modo, l'inviare un'informazione identica a più sistemi di recettori è lo stesso che inviarla a uno solo, cosa che ci consente di vedere come risulterebbe l'immagine se fosse possibile isolare un sistema di recettori.

A pagina 53 il lettore può vedere tre immagini in bianco e nero riprese attraverso filtri retinex che simulano la risposta spettrale dei pigmenti dei tre sistemi di coni. Le fragole e i ravanelli, per esempio, sono chiari nella ripresa a grandi lunghezze d'onda, più scuri in quella a lunghezze d'onda intermedie e più scuri ancora in quella a lunghezze minori. Benché l'arancia e il limone risultino altrettanto scuri delle fragole e dei ravanelli nell'immagine ripresa nella banda di lunghezze d'onda minori, presentano praticamente la medesima luminosità nelle altre due riprese. Sulla pagina stampata queste differenze sono assai piccole, eppure sono sufficienti all'occhio che vede una scena a colori per ottenere tutte le informazioni necessarie a distinguere innumerevoli sfumature di tutti i colori.

Quando le tre luminosità di una superficie sono state determinate da tre sistemi retinex, non sono necessarie altre informazioni per definire il colore di qualsiasi oggetto che sia presente nel campo visivo. Ogni specifico colore è così definito da una terna di determinate luminosità. A ogni terna di luminosità, inversamente, corrisponde un solo e ben determinato colore.

**La sensazione di luminosità è riportata in ascissa in scala lineare. Agli osservatori vengono mostrati un foglio di carta bianca (*9*) e uno di carta nera (*1*) e viene loro chiesto di scegliere un grigio di luminosità esattamente intermedia, denominato *5*. Allo stesso modo sono stati scelti i punti a mezzavia fra *1* e *5* e fra *5* e *9* e così via fino al completamento della scala. I punti estremi *0* e *10* sono determinati per estrapolazione. Infine si misurano le riflettanze integrate dei vari fogli di carta e si costruisce la curva che correla le riflettanze integrate nell'equivalente psicofisico.**

Le limitazioni della fotografia a colori rendono impossibile mostrare ai lettori gli esperimenti compiuti nel nostro laboratorio, che dimostrano senza ombra di dubbio l'indipendenza del colore percepito dal flusso di radiazioni che giunge all'occhio. I lettori potrebbero vedere due tavole quadrate di circa 1,35 metri di lato, ricoperte in maniera identica con circa 100 pezzi di carta di varia forma e colore. Per minimizzare gli effetti della riflessione speculare i pezzi di carta hanno una superficie opaca e, fatta eccezione per quelli neri, riflettono almeno il 10 per cento della radiazione incidente in qualsiasi zona dello spettro. In questi collage, che chiamiamo «Mondrian» (per la somiglianza con i quadri del noto pittore olandese), i pezzi di carta sono sistemati in modo che ognuno di essi sia circondato da almeno cinque o sei altri di colore diverso (*si veda la figura in alto a pagina 55*).

Ognuno dei due collage è illuminato da tre diversi proiettori muniti di filtri a banda passante molto stretta (non filtri retinex): le bande sono centrate sulle lunghezze d'onda di 670, 540 e 450 nanometri. Il flusso di luce di ognuno dei proiettori viene regolato da un trasformatore variabile e la durata dell'illuminazione è controllata con otturatori elettromagnetici sincronizzati. Infine v'è un fotometro a piccolo angolo di campo che può venire diretto su un punto qualsiasi dei due collage per misurare la radiazione riflessa e quindi il flusso che giunge all'occhio dell'osservatore. La lettura del fotometro viene proiettata su una scala posta sopra ai collage, dove può essere vista da chi partecipa all'esperimento.

Questo inizia illuminando il collage di sinistra, mentre quello di destra rimane al buio. I trasformatori variabili vengono posizionati in modo che tutti i pezzi del collage appaiano intensamente colorati, pur rimanendo ben bianchi quelli di tale colore. Questa sistemazione iniziale non è critica. Successivamente, accendendo un proiettore alla volta, si misura con il fotometro l'energia che raggiunge l'occhio da un pezzo di carta, per esempio un rettangolo bianco. Le letture della superficie bianca (in milliwatt per steradiante per metro quadrato) sono 65 unità per la luce di grande lunghezza d'onda, 30 per quella a lunghezza d'onda intermedia e 5 per quella di piccola lunghezza d'onda. Si sono stabilite in questo modo le energie associate con quella particolare sensazione di bianco.

Si spengono ora i tre proiettori che illuminano il collage di sinistra e si illumina quello a destra col solo proiettore a grande lunghezza d'onda, aggiustando il trasformatore variabile in modo che da un determinato pezzo di carta di colore qualsiasi giunga all'occhio un'energia uguale a quella che appena prima giungeva dal rettangolo bianco del collage di sinistra, 65 unità. Spegnamo il proiettore e aggiustiamo separatamente gli altri due, uno dopo l'altro, in modo che anche con questi l'energia riflessa da quel pezzo di carta sia la stessa che prima proveniva dal rettangolo bianco del collage di sinistra. Non abbiamo ancora acceso contemporaneamente i tre proiettori, ma sappiamo che quando lo faremo la tripletta di energie che verrà riflessa dalla superficie, di colore ancora sconosciuto, scelta sul secondo collage sarà identica a quella che, sul primo collage, forniva una sensazione di bianco.

Quando accendiamo tutti e tre i proiettori scopriamo invece, per esempio, che il pezzo del collage di destra scelto per l'esperimento è verde. Illuminiamo anche il collage di sinistra con i proiettori regolati come in precedenza, in modo da consentire la visione contemporanea dei due collage. Il rettangolo bianco a sinistra continua a essere bianco, mentre la carta verde a destra continua a essere verde. Eppure entrambe le superfici riflettono verso l'osservatore la medesima tripletta di energie: 65, 30 e 5 nelle unità da noi scelte.

Ora spegnamo tutti i riflettori, scegliamo qualche altro pezzo del collage di sinistra e aggiustiamo i proiettori in modo che anche questo rifletta la medesima tripletta di energie che ha dato origine alla sensazione di bianco sul collage di sinistra e di verde su quello di destra. Riaccendendo contemporaneamente i tre proiettori che illuminano il collage di sinistra scopriamo, per esempio, che il nuovo pezzo di carta è giallo e, ancora una volta, possiamo vedere contemporaneamente le superfici gialla e verde, la prima a sinistra e la seconda a destra.

È possibile continuare l'esperimento scegliendo altre zone dei due collage, di colore blu, grigio, rosso e così via. Il risultato dimostra che la sensazione del colore non è affatto correlata al prodotto del coefficiente di riflessione per l'illuminazione, cioè all'energia riflessa, per quanto questa informazione sembri essere la sola che giunge all'osservatore dalle varie zone dei collage.

Per dimostrare che negli esperimenti appena descritti la sensazione del colore non dipende da un estensivo adattamento cromatico dei pigmenti della retina, i proiettori sono muniti di otturatori sincronizzati, in modo da consentire la visione dei collage solo per un tempo molto breve, un decimo di secondo o meno. I risultati dell'esperimento non cambiano, qualunque sia la durata dell'osservazione. È quindi possibile dire che né l'adattamento cromatico né i movimenti degli occhi sono coinvolti nei fenomeni che producono la sensazione del colore. Infine il disegno dei collage minimizza l'importanza della forma e delle dimensioni delle superfici circostanti, della familiarità degli oggetti e della memoria dei colori. Di quando in quando v'è qualche tentativo casuale di spiegare i risultati degli esperimenti descritti in base a quella che viene detta costanza dei colori. È chiaro che parlare di costanza dei colori è solo un modo conciso per designare quella notevole abilità dell'occhio che costituisce l'argomento del presente articolo.

La cosa misteriosa è come tutti possiamo accordarci con precisione nella definizione dei colori che vediamo quando non esiste alcuna grandezza fisica apparente che ci consenta di specificare il colore di un oggetto. In effetti si può dire che lo stimolo che produce la sensazione di colore di una superficie non è la radiazione riflessa da questa. È compito della psicofisica trovare lo stimolo che produce la sensazione di colore.

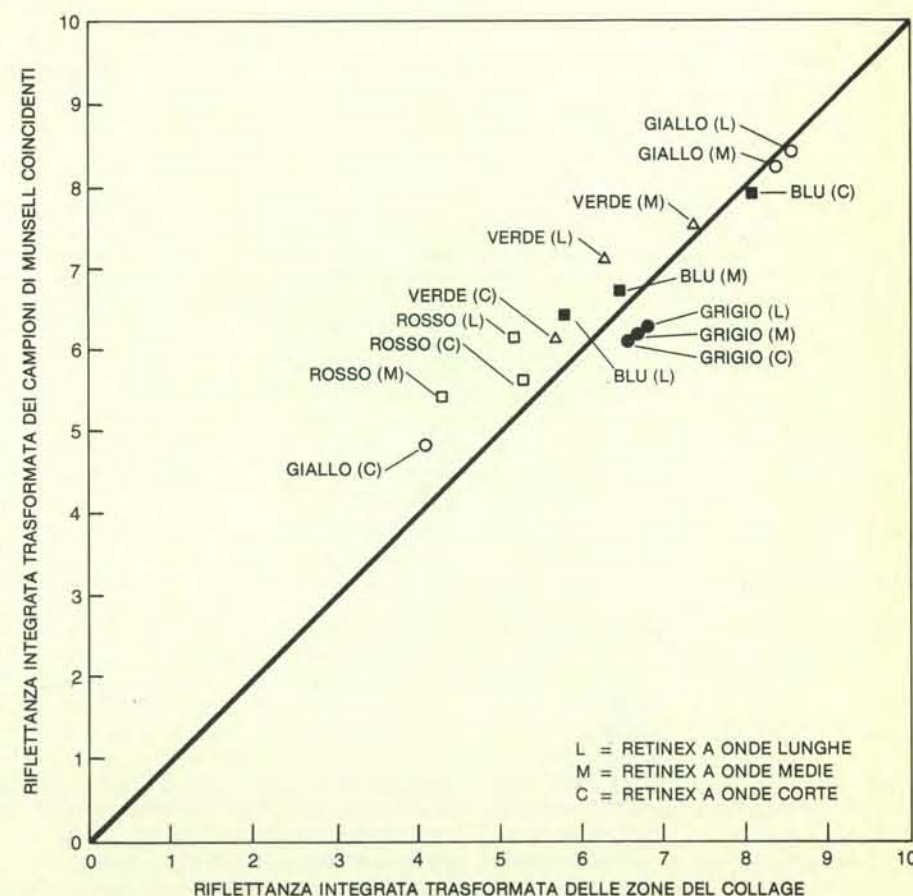
Ricordiamo che ciò che l'occhio fa, senza mai sbagliare, è scoprire l'entità della luminosità, indipendentemente dal flusso. Abbiamo visto che questo è vero per un solo sistema di recettori, i bastoncelli, quando opera singolarmente, e per i tre sistemi di coni quando operano collettivamente nell'osservazione di superfici bianche, grige e nere. Illuminiamo ora i collage colorati con la luce proveniente da uno solo dei proiettori, per esempio quello che fornisce una banda di radiazioni di grande lunghezza d'onda, e osserviamo cosa succede quando si aumenta e diminuisce di molto l'illuminazione. Si osserva che le varie zone del collage mantengono una scala di luminosità costante. Ma se cambiamo il proiettore e usiamo, per esempio, quello con filtro a banda passante intermedia, la luminosità di molti dei pezzi di carta cambia: molti fra i circa cento pezzi di carta vengono a occupare un posto diverso nella scala delle luminosità. Usando il terzo proiettore si troverà una terza successione di luminosità. In particolare, una superficie rossa verrà vista chiara con luce a grande lunghezza d'onda, più scura in luce a media lunghezza d'onda e molto scura in quella a corta lunghezza d'onda. D'altro canto una superficie blu apparirà chiara in luce a corta lunghezza d'onda e molto scura in luce a media e grande lunghezza d'onda. Superfici di altro colore mostreranno altre triplette di luminosità. Quando abbiamo fatto esperimenti di questo tipo, quasi vent'anni fa, siamo stati portati inevitabilmente a concludere che le triplette di luminosità forniscono quello stimolo costante e indipendente dal flusso di cui abbiamo bisogno per provare la sensazione del colore.

È evidente che le luminosità di un certo pezzo di carta colorata illuminato con sorgenti in tre diverse bande dello spettro sono correlate con la quantità di energia che esso riflette alle diverse lunghezze d'onda. Vediamo ora, con un esperimento particolare, come sia possibile correlare tali riflessioni con le luminosità percepite e come, nel corso di questo processo, il flusso di energia che raggiunge l'occhio diventi irrilevante per la percezione del colore.

Nel nostro laboratorio McCann, Suzanne P. McKee e Thomas H. Taylor hanno condotto uno studio sistematico delle risposte degli osservatori di un collage semplificato, con superfici di 17 colori diversi. Essi chiedevano agli osservatori di confrontare le 17 superfici, una alla volta, e sotto illuminazioni diverse, con degli standard quadrati di carta colorata scelti da un campionario di colori, il *Munsell Book of Color*, che venivano invece osservati sotto un'illuminazione «bianca» costante.

Le sorgenti che illuminavano il collage venivano aggiustate in cinque esperimenti diversi in modo che cinque differenti pezzi del collage (di colore grigio, rosso, giallo, blu e verde) riflettessero la medesima tripletta di flussi di radiazione. Lo osservatore cominciava col scegliere nella serie di standard i colori che gli parevano uguali a quelli dei 17 pezzi del collage quando il pezzo grigio rifletteva una particolare tripletta di energie. Successivamente una seconda serie di colori standard veniva selezionata sotto un'illuminazione tale per cui un pezzo rosso del collage rifletteva la medesima tripletta di energie prima riflessa dal pezzo grigio. L'operazione veniva poi ripetuta con il collage illuminato in modo da far riflettere sempre la stessa tripletta ai pezzi di carta giallo, blu e verde.

Le figure a pagina 57 mostrano i dettagli dell'esperimento e i cinque diversi colori standard di Munsell che gli osservatori hanno selezionato come uguali a quelli delle cinque superfici studiate quando ognuna di esse rifletteva esattamente la medesima tripletta di energie. Malgrado le energie riflesse dal collage che raggiungevano uno degli occhi fossero sempre le stesse, tutti gli osservatori, usando il secondo occhio, sceglievano colori standard grigi, rossi, gialli, blu e verdi.

L'illuminazione costante usata per la

**L'accordo fra le riflettanze integrate trasformate delle zone del collage e dei colori standard scelti come identici è riassunto in questa figura per tutte e tre le bande spettrali. Le riflettanze integrate trasformate delle cinque zone del collage e degli identici colori standard sono state determinate come descritto nella figura di pagina 59. Qui le triplette delle riflettanze integrate trasformate delle zone usate per riflettere identiche triplette di energia sono riportate in funzione di quelle a fronte dei colori standard scelti come identici. I punti nel loro insieme rappresentano la corrispondenza per le tre bande spettrali, ma ognuno di essi rappresenta la correlazione fra un colore del collage e il corrispondente standard in una banda spettrale. La linearità della correlazione dimostra che la riflettanza integrata trasformata è il corrispondente fisico della sensazione di «luminosità» e mostra la precisione con cui una tripletta di luminosità determina il colore.**

| ILLUMINAZIONE UNIFORME (UNITÀ RELATIVE) | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | 8 | 8 | 8 | 8 | 8 | 8 | 100 | 100 | 50 | 50 | 50 | 50 | 72 | 72 | 72 | 55 | 55 | 55 |
| 100 | 8 | 8 | 8 | 8 | 8 | 8 | 100 | 100 | 50 | 50 | 50 | 50 | 72 | 72 | 72 | 55 | 55 | 55 |
| 100 | 8 | 8 | 8 | 8 | 8 | 8 | 100 | 100 | 50 | 50 | 50 | 50 | 72 | 72 | 72 | 55 | 55 | 55 |
| 100 | 23 | 23 | 23 | 38 | 38 | 38 | 100 | 100 | 50 | 50 | 50 | 50 | 72 | 72 | 72 | 55 | 55 | 55 |
| 100 | 23 | 23 | 85 | 85 | 85 | 38 | 100 | 100 | 50 | 50 | 50 | 43 | 43 | 72 | 72 | 55 | 55 | 55 |
| 100 | 23 | 23 | 85 | 85 | 85 | 38 | 100 | 100 | 36 | 36 | 36 | 43 | 43 | 72 | 72 | 55 | 55 | 55 |
| 100 | 23 | 23 | 23 | 38 | 38 | 38 | 100 | 100 | 36 | 36 | 36 | 43 | 43 | 72 | 72 | 55 | 55 | 55 |
| 100 | 23 | 23 | 23 | 38 | 38 | 2 | 2 | 2 | 2 | 73 | 73 | 43 | 43 | 72 | 72 | 55 | 55 | 55 |
| 100 | 23 | 23 | 23 | 38 | 38 | 2 | 2 | 2 | 2 | 73 | 73 | 43 | 43 | 72 | 72 | 55 | 55 | 55 |
| 100 | 23 | 23 | 23 | 38 | 38 | 2 | 2 | 2 | 2 | 73 | 73 | 73 | 73 | 50 | 50 | 50 | 50 | 50 |
| 100 | 23 | 23 | 23 | 38 | 38 | 2 | 2 | 2 | 2 | 73 | 73 | 73 | 73 | 50 | 50 | 50 | 50 | 50 |
| 100 | 23 | 23 | 23 | 38 | 38 | 2 | 2 | 2 | 2 | 73 | 73 | 73 | 73 | 50 | 50 | 50 | 50 | 50 |
| 100 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 73 | 73 | 50 | 50 | 50 | 50 | 50 |
| 100 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 73 | 73 | 73 | 18 | 18 | 18 | 18 |
| 100 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 73 | 73 | 73 | 18 | 18 | 18 | 18 |
| 100 | 41 | 41 | 41 | 41 | 41 | 41 | 41 | 43 | 43 | 43 | 43 | 43 | 43 | 18 | 18 | 18 | 18 | 18 |
| 100 | 41 | 41 | 41 | 41 | 41 | 41 | 41 | 43 | 43 | 43 | 43 | 43 | 43 | 18 | 18 | 18 | 18 | 18 |
| 100 | 41 | 41 | 41 | 41 | 41 | 41 | 41 | 43 | 43 | 43 | 43 | 43 | 43 | 18 | 18 | 18 | 18 | 18 |

| FLUSSO DI LUCE | RAPPORTO DOPO IL TEST DI UGUAGLIANZA | SEQUENZA DEI PRODOTTI |
|---|---|---|
| 100 / 100 | 1,0 | 1,0 |
| 100 / 2 | 0,02 | 0,02 |
| 2 / 73 | 36,5 | 0,73 |
| 73 / 18 | 0,25 | 0,18 |
| 18 / 18 | 1,0 | 0,18 |

**Rimane da stabilire il metodo seguito dall'occhio per definire la luminosità nelle immagini complesse. In questa figura e in quella della pagina a fronte è descritto un metodo efficace e fisiologicamente possibile. I numeri entro lo schema del collage rappresentano le radianze integrate nella banda a grande lunghezza d'onda riflesse da ognuna delle zone quando l'insieme viene illuminato uniformemente. Il sistema retinex sensibile alle grandi lunghezze d'onda «misura» indipendentemente le radianze integrate in tale banda, punto per punto lungo un percorso arbitrario (*in colore*). I valori dei flussi in ogni coppia di punti adiacenti vengono convertiti in rapporto. Questi rapporti vengono sottoposti a una verifica di uguaglianza: si assume che esista una variazione quando un rapporto differisce da quello immediatamente precedente di più di un valore soglia (che nei nostri calcoli è stato assunto 0,003). Se il rapporto non si scosta dall'unità entro tali limiti, esso viene ritenuto invariato e posto uguale a uno. Quando lungo il cammino il rapporto si scosta dall'unità, esso viene moltiplicato per quello precedente. Questo prodotto è contemporaneamente la risposta del sistema per quella variazione di riflettanza integrata e il segnale da trasmettere perché venga moltiplicato con il prossimo rapporto diverso da uno. Quando il cammino attraversa una linea di separazione fra zone di diversa luminosità, si ha un brusco cambiamento del rapporto fra radianze integrate e quindi anche nella sequenza dei prodotti. Qui il cammino inizia nella zona bianca, dove il flusso di energia raggiante è 100. Appena si giunge alla zona marrone in basso a destra il prodotto diviene 0,18. Il sistema retinex ha quindi determinato che l'area marrone riflette solo il 18 per cento delle radiazioni di grande lunghezza d'onda rispetto alla zona bianca. Qualsiasi altro cammino che terminasse nella zona marrone attraversando in qualche momento quella bianca avrebbe dato il medesimo risultato. Mediando le risposte per le varie superfici ottenute seguendo diversi cammini arbitrari, il sistema retinex sensibile alle grandi lunghezze d'onda giunge a stabilire per ogni zona di colore un unico valore di riflettanza, che definisce la luminosità percepita. Contemporaneamente pure i sistemi retinex sensibili a lunghezze d'onda corte e a lunghezze d'onda intermedie stabiliscono i loro valori di luminosità. Il confronto delle triplette di luminosità per ogni superficie fornisce la sensazione del colore.**

osservazione degli standard di Munsell era costituita da una tripletta di sorgenti in tre bande dello spettro aggiustata in modo che gli osservatori stessi avevano giudicato produrre il «miglior bianco». Il flusso di energia riflesso dal campione più bianco fra gli standard di Munsell era costituito da 11,5 unità a grande lunghezza d'onda, 7,8 a lunghezza d'onda intermedia e 3,3 a piccola lunghezza d'onda. Le sorgenti fornivano bande spettrali strette centrate a 630, 530 e 450 nanometri. Il collage veniva illuminato da tre sorgenti identiche, con le radiazioni mescolate in proporzioni variabili.

A questo punto ci si potrebbe porre una domanda: una superficie grigia non potrebbe mostrare una marcata variazione di colore quando i pezzi di carta circostanti riflettono luci di composizione spettrale molto diversa? E simili cambiamenti di colore non potrebbero render conto dei risultati degli esperimenti con i collage? La risposta è che nessuna manipolazione dei pezzi di carta circostanti riesce a far coincidere il foglio grigio con gli standard di Munsell rosso, giallo, blu e verde scelti dagli osservatori negli esperimenti con il collage.

McCann, John A. Hall e io abbiamo ulteriormente approfondito la questione ripetendo l'esperimento di confronto del collage con gli standard di Munsell in varie condizioni, in modo che la composizione spettrale media della luce riflessa dal collage e dai suoi dintorni fosse la stessa, qualunque fosse la composizione spettrale necessaria per ottenere la riflessione di una tripletta costante da varie zone di esso. Questo è stato ottenuto in un caso circondando il collage con i pezzi di carta vivacemente colorati e scelti in modo che compensassero esattamente la composizione spettrale media della luce riflessa dal collage. In un'altra condizione sono stati tagliati i 17 pezzi di carta che costituivano il collage, disponendoli poi ben distanziati su sfondi di colore contrastante. Nessuna delle due disposizioni modifica in maniera significativa la scelta dei colori standard identici a quelli delle varie zone del collage.

Torniamo allora a cercare lo stimolo che ci guida in maniera tanto accurata alla giusta identificazione dei colori. Se non si tratta di un flusso di energia raggiante che giunge all'occhio dai vari punti del campo visivo, quali sono i corrispettivi fisici delle luminosità degli oggetti in tre diverse bande dello spettro, corrispondenti alle sensibilità spettrali dei pigmenti dei coni? Può venire messo in evidenza un tale corrispettivo fisico della sensazione di luminosità?

McCann, McKee e Taylor proseguirono il lavoro misurando la radianza, cioè l'energia riflessa verso l'osservatore, delle varie zone del collage e degli standard di colore utilizzando un fotomoltiplicatore equipaggiato con filtri retinex. Dato che il complesso fotomoltiplicatore-filtro retinex integra il flusso di energia raggiante in una ampia banda di lunghezze d'onda, lo strumento fornisce un valore di radianza che chiameremo radianza integrata. McCann e i suoi colleghi hanno quindi ottenuto le radianze integrate di un grosso foglio di carta bianca illuminato con le combinazioni di luci usate per illuminare il collage nei vari esperimenti di confronto. Se facciamo il rapporto fra la radianza integrata di una zona del collage e quella misurata sulla carta bianca, si ottiene un valore di riflettanza integrata, esprimibile in forma percentuale.



Le riflettanze integrate dei vari colori standard di Munsell sono state poi determinate nello stesso modo, sotto l'illuminazione «bianca» costante. È stata cioè misurata la riflettanza percentuale impiegando detettori aventi la medesima sensibilità spettrale dei pigmenti visivi. I risultati mostrano che i colori standard scelti come identici a quelli delle varie zone del collage possiedono approssimativamente le medesime riflettanze integrate percentuali delle zone suddette. La zona blu del collage, per esempio, possiede una tripletta di riflettanze integrate nelle bande a lunga, media e corta lunghezza d'onda pari a 27,3, 35,9 e 60,7 per cento. I corrispettivi valori dello standard Munsell sono 34,6, 38,5 e 57,1 per cento (*si veda l'illustrazione a pagina 59*).

Infine le riflettanze integrate possono venire disposte in successione in modo che una loro spaziatura uniforme corrisponda a una sensazione di uguale differenziazione delle luminosità. La curva che descrive la legge di questa trasformazione è mostrata nell'illustrazione alla pagina precedente. Utilizzando tale curva si osserva che la zona blu del collage ha una tripletta di riflettanze integrate trasformate di 5,8, 6,5 e 8,1, mentre il corrispondente colore standard possiede i valori 6,4, 6,7 e 7,9. Se consideriamo le cinque zone del collage che hanno inviato all'osservatore identiche triplette di energia e confrontiamo le loro riflettanze integrate trasformate con quelle dei corrispondenti standard di Munsell, troviamo che tutti i valori sono in ottimo accordo. In altre parole abbiamo trovato nelle triplette di riflettanze integrate un accurato corrispettivo fisico della sensazione di colore. I dati cadono lungo la retta che definisce la correlazione perfetta, che è la bisettrice degli assi (*si veda l'illustrazione a pagina 61*).

Abbiamo cercato un corrispettivo fisico della sensazione di luminosità e abbiamo trovato che le riflettanze integrate trasformate delle cinque zone di colore che hanno riflesso uguali triplette di energie sono le stesse di quelle dei colori standard uguali alle zone suddette. Questa correlazione ci consente di utilizzare le riflettanze integrate trasformate come una misura indiretta della luminosità. Il problema ora diviene quello di trovare come fa l'occhio a determinare la luminosità che corrisponde alle riflettanze degli oggetti nelle varie bande dello spettro.

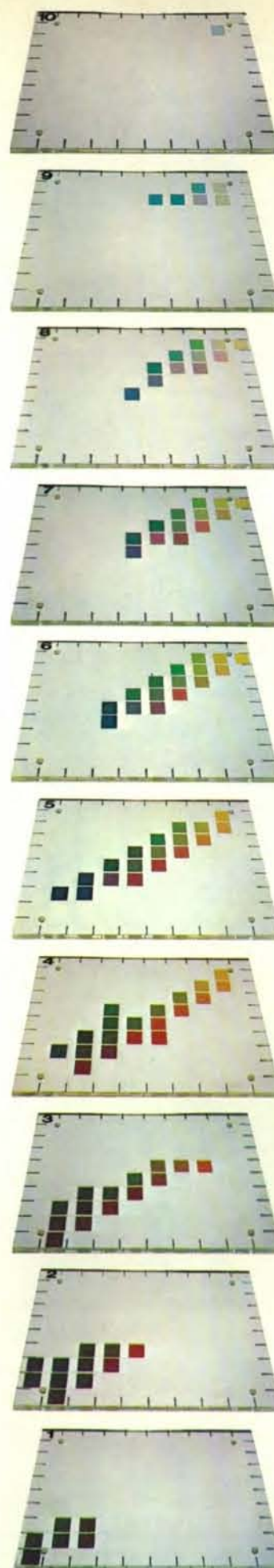
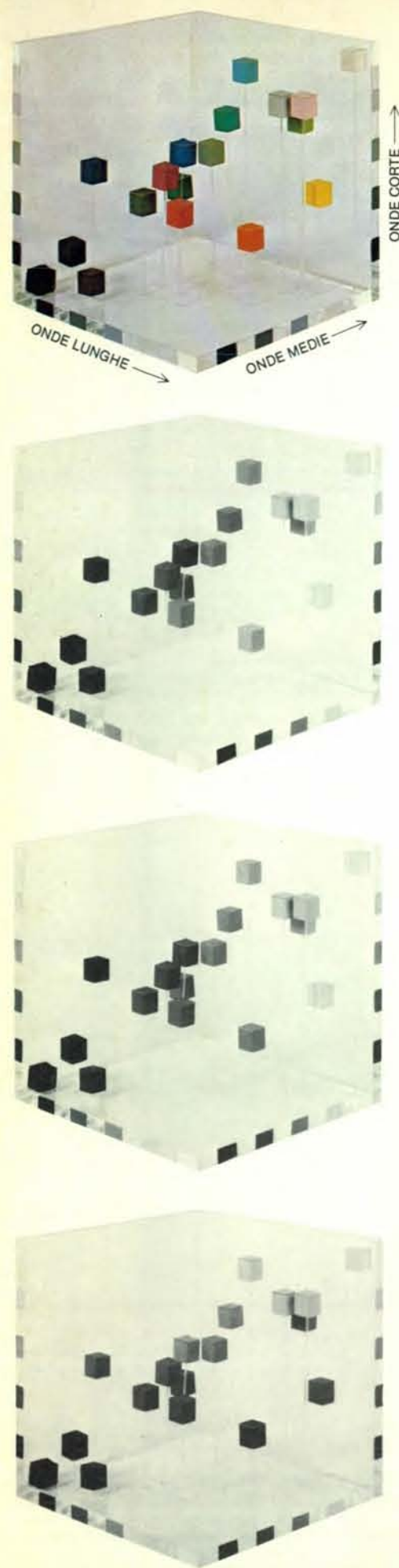
Una cosa è misurare la tripletta di equivalenti di luminosità impiegando un fotomoltiplicatore accoppiato a un filtro retinex, e una cosa ben diversa è, per l'occhio, determinare le luminosità in un mondo illuminato in maniera non uniforme e privo di pezzi di carta bianca da confrontare. È stata descritta in precedenza la capacità di un sistema isolato di recettori - quello ipersensibile dei bastoncelli - di classificare correttamente gli oggetti in funzione del loro potere intrinseco di riflessione, indipendentemente dal fatto che si trovino in zone poco o molto illuminate del campo visivo. Il fatto che un sistema fotorecettore abbia una simile capacità rende plausibile l'ipotesi che anche gli altri tre sistemi, destinati alla visione diurna, la possiedano. Dato che questi tre sistemi vedono il mondo in tre ampie, ma comunque delimitate zone dello spettro (che sono quelle utilizzate nei filtri retinex), ognuno di essi forma un'immagine di luminosità distinta dalle altre. Queste immagini non vengono mescolate, bensì confrontate. È appunto il confronto delle tre luminosità di una determinata superficie che fornisce la sensazione del colore.

Come fa un sistema biologico a costruire una gerarchia e una distribuzione spaziale di luminosità in una scena senza sapere nulla né sul tipo di illuminazione né sui poteri riflettenti dei vari oggetti? Il meccanismo che descriverò ora è il più generale da noi trovato in grado di sormontare tali difficoltà.

Vorrei segnalare innanzitutto l'importanza dei margini nella definizione degli oggetti e delle superfici presenti in una scena. Se illuminiamo un foglio di carta bianca con una sorgente posta di lato, non si osserva alcuna discontinuità di colore. Immaginiamo ora di misurare la luminanza del foglio con due detettori puntati verso punti diversi del foglio. Se l'illuminazione non è uniforme le luminanze dei due punti risulteranno ovviamente diverse, ma avvicinando i detettori i valori di luminanza tenderanno ad avvicinarsi e il loro rapporto si approssimerà all'unità. Se però i due detettori si trovano da parti opposte rispetto al con-

| ILLUMINAZIONE NON UNIFORME (UNITÀ RELATIVE) | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 37 | 3 | 3 | 3 | 3 | 3 | 3 | 37 | 37 | 18 | 18 | 18 | 18 | 27 | 27 | 27 | 21 | 21 | 21 |
| 43 | 3 | 3 | 3 | 3 | 3 | 3 | 43 | 43 | 21 | 21 | 21 | 21 | 31 | 31 | 31 | 24 | 24 | 24 |
| 49 | 4 | 4 | 4 | 4 | 4 | 4 | 49 | 49 | 25 | 25 | 25 | 25 | 35 | 35 | 35 | 27 | 27 | 27 |
| 57 | 13 | 13 | 13 | 22 | 22 | 22 | 57 | 57 | 28 | 28 | 28 | 28 | 41 | 41 | 41 | 31 | 31 | 31 |
| 65 | 15 | 15 | 55 | 55 | 55 | 25 | 65 | 65 | 33 | 33 | 33 | 28 | 28 | 47 | 47 | 36 | 36 | 36 |
| 76 | 18 | 18 | 64 | 64 | 64 | 29 | 76 | 76 | 27 | 27 | 27 | 33 | 33 | 54 | 54 | 42 | 42 | 42 |
| 87 | 20 | 20 | 20 | 33 | 33 | 33 | 87 | 87 | 32 | 32 | 32 | 38 | 38 | 63 | 63 | 48 | 48 | 48 |
| 100 | 23 | 23 | 23 | 38 | 38 | 2 | 2 | 2 | 2 | 74 | 74 | 44 | 44 | 72 | 72 | 56 | 56 | 56 |
| 116 | 27 | 27 | 27 | 44 | 44 | 2 | 2 | 2 | 2 | 85 | 85 | 50 | 50 | 83 | 83 | 64 | 64 | 64 |
| 133 | 31 | 31 | 31 | 51 | 51 | 2 | 2 | 2 | 2 | 98 | 98 | 98 | 98 | 66 | 66 | 66 | 66 | 66 |
| 154 | 36 | 36 | 36 | 59 | 59 | 3 | 3 | 3 | 3 | 113 | 113 | 113 | 113 | 76 | 76 | 76 | 76 | 76 |
| 178 | 42 | 42 | 42 | 68 | 68 | 3 | 3 | 3 | 3 | 131 | 131 | 131 | 131 | 88 | 88 | 88 | 88 | 88 |
| 205 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 32 | 151 | 151 | 102 | 102 | 102 | 102 | 102 |
| 236 | 37 | 37 | 37 | 37 | 37 | 37 | 37 | 37 | 37 | 37 | 37 | 174 | 174 | 174 | 43 | 43 | 43 | 43 |
| 272 | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 42 | 200 | 200 | 200 | 49 | 49 | 49 | 49 |
| 314 | 129 | 129 | 129 | 129 | 129 | 129 | 129 | 135 | 135 | 135 | 135 | 135 | 135 | 57 | 57 | 57 | 57 | 57 |
| 362 | 149 | 149 | 149 | 149 | 149 | 149 | 149 | 156 | 156 | 156 | 156 | 156 | 156 | 66 | 66 | 66 | 66 | 66 |
| 417 | 172 | 172 | 172 | 172 | 172 | 172 | 172 | 179 | 179 | 179 | 179 | 179 | 179 | 76 | 76 | 76 | 76 | 76 |

| FLUSSO DI LUCE | RAPPORTO DOPO IL TEST DI UGUAGLIANZA | SEQUENZA DEI PRODOTTI |
|---|---|---|
| 60,9 / 61,0 | 1,0 | 1,0 |
| 95,1 / 1,9 | 0,02 | 0,02 |
| 2,4 / 87,6 | 36,5 | 0,73 |
| 185,1 / 46,3 | 0,25 | 0,18 |
| 55,8 / 55,9 | 1,0 | 0,18 |

**Una situazione più realistica di quella della figura della pagina precedente con illuminazione non uniforme, viene risolta altrettanto bene con il metodo dei prodotti in sequenza. Si giunge infatti al valore 0,18 per la zona marrone al termine del cammino anche se in questo caso il sistema retinex sensibile alle lunghezze d'onda maggiori riceve un flusso identico (*57*) dal centro della zona marrone e dal centro di quella bianca. Il metodo consente quindi di giungere al calcolo di una riflettanza che non dipende dal flusso di energia. I valori esatti dei flussi di luce lungo il cammino sono stati ottenuti con un programma di calcolo per elaboratore che interpola altri 75 valori fra due adiacenti stampati su questo schema.**

fine fra due aree che differiscono bruscamente in riflettività, come nel caso, per esempio, di una superficie grigia posta sopra quella bianca, il rapporto fra i segnali in uscita dei due detettori si avvicinerà al rapporto fra le due riflettanze. Pertanto il semplice procedimento di fare il rapporto fra le luminanze di due punti adiacenti consente contemporaneamente di scoprire il confine fra due zone di diverso colore e di eliminare l'effetto di un'illuminazione non uniforme. Se la intera immagine viene elaborata in termini di rapporti di luminanza fra punti vicini, si ottiene una serie di numeri adimensionali che risultano indipendenti dall'illuminazione effettiva, e che forniscono il rapporto fra le riflettanze al confine fra superfici diverse; i valori delle riflettanze sono però ancora indeterminati.

Per determinare le riflettanze occorre correlare questi rapporti fra riflettanze in tutto il campo visivo. Dato il rapporto fra le luminanze ai margini di due superfici, che chiamiamo 1 e 2, possiamo moltiplicarlo per il rapporto delle luminanze al margine fra la seconda e una terza. Questo prodotto di rapporti approssima il rapporto fra le riflettanze della prima e della terza superficie, qualunque sia la distribuzione dell'illuminazione. Analogamente possiamo ottenere il rapporto fra le riflettanze di due superfici qualsiasi entro il campo visivo, per quanto distino fra di loro, moltiplicando i rapporti di riflettanza calcolati a tutti i confini fra la superficie di partenza e l'ultima. Possiamo anche stabilire il rapporto di riflettanza per qualsiasi zona troncando la sequenza dei prodotti (*si vedano le illustrazioni alle pagine 62 e 63*).

**Il «solido» dei colori mostra la disposizione di tutti i colori percepiti, compresi il bianco e il nero, in uno spazio tridimensionale strutturato secondo la teoria della visione dei colori proposta dall'autore. La posizione di un colore in questo spazio non è definita da una tripletta di energie riflesse, bensì dalla tripletta di luminosità calcolata dall'occhio. La fotografia a colori in alto mostra la disposizione spaziale di alcuni colori rappresentativi. La direzione dell'aumento delle luminosità lungo i tre assi è segnata dalle frecce. Le tre fotografie in bianco e nero del solido dei colori sono state prese con combinazioni di filtri e pellicole retinex. Esse mostrano le luminosità dei colori come verrebbero percepite separatamente dai pigmenti visivi sensibili a lunghezze d'onda grandi (*in alto*), intermedie (*al centro*) e corte (*in basso*). Le dieci immagini a colori sulla destra rappresentano sezioni orizzontali del solido dei colori. Ogni piano è il luogo dei colori possibili con una luminosità costante nella banda delle lunghezze d'onda minori. Per esempio, il quinto piano dal basso mostra la varietà delle sensazioni di colore ottenibili con tutti i valori possibili di luminosità alle lunghezze d'onda medie e lunghe quando queste vengono combinate con un valore di luminosità 5 nella banda delle lunghezze d'onda inferiori. I quadrati colorati sono campioni presi dal *Munsell Book of Color*. In generale le zone vuote nei singoli piani indicano regioni in cui il colore potrebbe venire prodotto solo con pigmenti fluorescenti, ammesso che sia possibile.**



Stiamo avvicinandoci alla risposta alla domanda fondamentale: come si riesce a determinare la riflettanza di una superficie senza porre vicino a questa uno standard di confronto? La sequenza di prodotti può venire utilizzata in modo da evitare di dover porre due superfici una vicina all'altra, fornendo quindi un equivalente di operazione fotometrica fattibile per l'occhio.

Rimane quindi da definire come l'occhio possa scoprire la superficie di maggior riflettanza entro un campo visivo e decidere se si tratti di un bianco o di un altro colore. Nel modello che abbiamo proposto le sequenze di prodotti vengono calcolate lungo svariati cammini arbitrari che si muovono attraverso la disposizione bidimensionale di energie sulla «retina» modello. Dato che questi cammini possono iniziare in un punto qualsiasi, e non proprio nel punto di maggior riflettanza, il primo valore di ogni sequenza viene assunto arbitrariamente come 100 per cento. Ne segue che la sequenza di prodotti diviene maggiore dell'unità ogni volta che il cammino attraversa una zona di riflettanza maggiore di quella del punto iniziale.

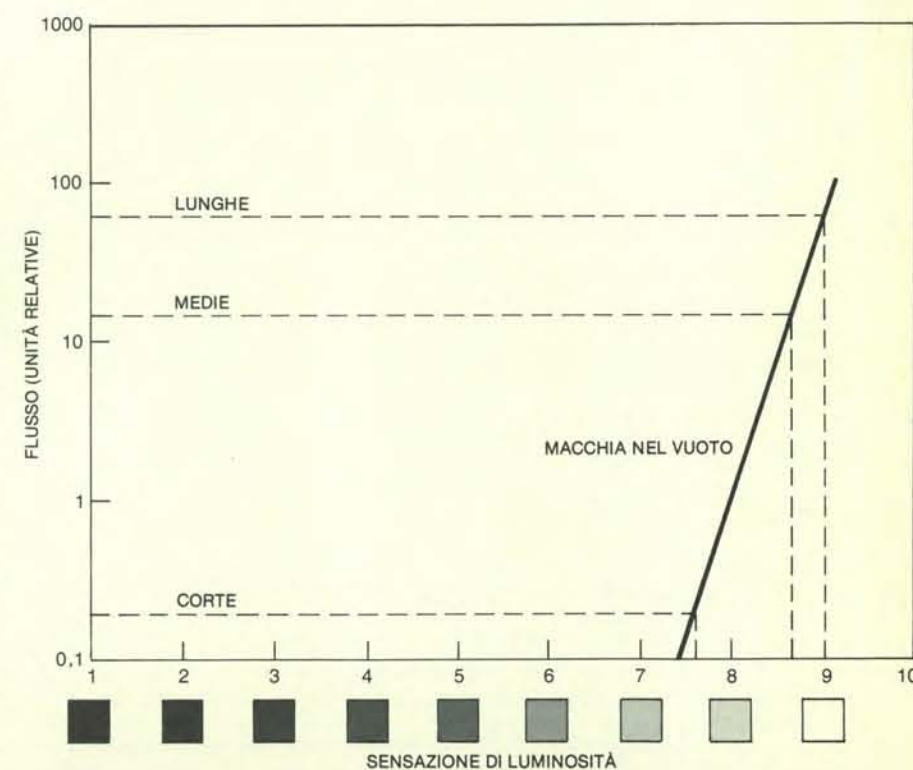
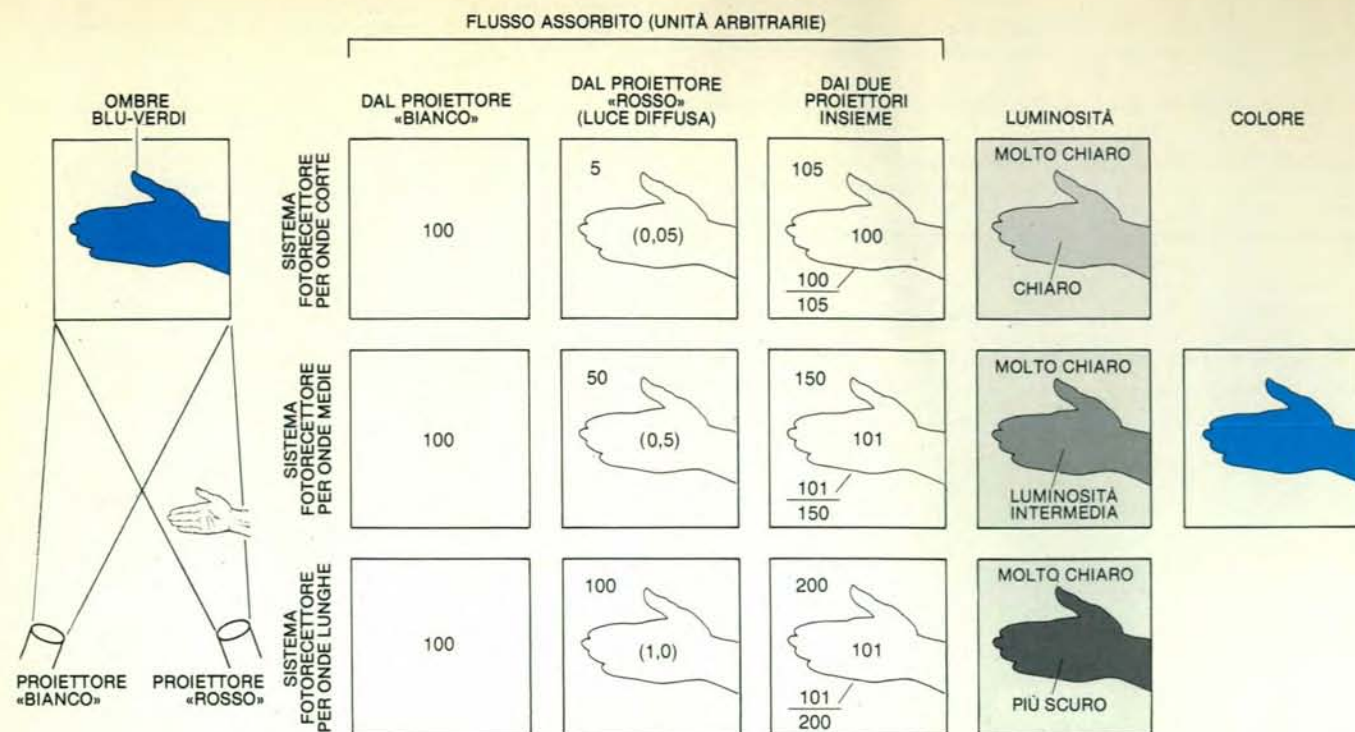
Il raggiungimento di un valore superiore all'unità nella sequenza di prodotti indica che la sequenza deve ricominciare da capo assumendo l'ultimo punto di maggior riflettanza come nuovo 100 per cento. Questo procedimento consente il reperimento della zona di maggior riflettanza lungo il cammino percorso dall'occhio. Una volta trovato questo punto tutti i prodotti calcolati successivamente diventano funzione di tale maggior riflettanza. È stato possibile programmare in maniera soddisfacente un calcolatore per studiare il numero dei percorsi, le loro lunghezze e convoluzioni, i valori di soglia per il riconoscimento dei margini e, forse la cosa più importante, per studiare come utilizzare tutti i percorsi che si originano nei diversi punti del campo visivo.

Il sistema biologico utilizza un programma di questo genere in qualche parte non ben definita del percorso fra la retina e la corteccia cerebrale. Il processo corrispondente al calcolo delle successioni di prodotti non implica la valutazione di valori medi né di superfici né di flussi di energia. Richiede comunque l'esecuzione di operazioni aritmetiche su tutto il campo visivo. Inoltre, dato che ciò che interessa viene visto per un tempo molto breve, tutti i calcoli e le conclusioni relative ai rapporti di luminosità devono venir effettuati in frazioni di secondo e in maniera indipendente dai movimenti dell'occhio. Con un singolo impulso di luce, il movimento dell'occhio, per definizione, non è necessario. Con un'illuminazione continua, i normali movimenti rapidi dell'occhio servono probabilmente per effettuare l'aggiornamento dei calcoli.

Con un calcolatore programmato secondo il nostro modello possiamo ottenere, per ogni zona del collage a colori, triplette di luminosità che corrispondono ottimamente alle luminosità che si potrebbero misurare con un fotomoltiplicatore posto dietro un filtro retinex. Il colore corrispondente a ogni tripletta può essere visualizzato con l'aiuto del «solido dei colori» che abbiamo realizzato, nel quale i colori di Munsell sono disposti lungo le tre dimensioni dello spazio in funzione delle luminosità misurate in tre bande dello spettro con filtri retinex (*si veda la figura nella pagina a fronte*).

Normalmente la sensazione di luce bianca verrà prodotta da qualsiasi superficie che venga posta al vertice della scala di luminosità da tutti e tre i sistemi retinex. D'altro canto una superficie che si trovi al vertice di solo due delle tre scale di luminosità verrà vista come un altro colore. Pertanto una superficie che si trovi al vertice delle scale di luminosità dei sistemi recettori per grandi e medie lunghezze d'onda ma sia meno luminosa di altre superfici per il sistema sensibile alle piccole lunghezze d'onda non verrà vista bianca bensì gialla. Un simile confronto di triplette di luminosità dei vari punti di una medesima scena procura la sensazione di colore, zona per zona, malgrado l'esistenza di imprevedibili variazioni dell'illuminazione.

Osservando le fotografie in bianco e nero prese attraverso filtri retinex si nota una grossa differenza fra le luminosità della maggior parte degli oggetti nelle immagini che rappresentano il sistema a piccola lunghezza d'onda e in quelle che rappresentano gli altri due sistemi. Eppure sono proprio le piccole differenze di luminosità registrate dai sistemi sensibili alle medie e grandi lunghezze d'onda che sono responsabili della percezione dei verdi e rossi più vivaci.

Simili risposte, sensibili e accurate a piccole differenze di luminosità, stanno alla base della visione dei colori in condizioni di illuminazione anomale, molto lontane da quelle per cui si è evoluto l'occhio. Due esempi interessanti riguardano il colore di una macchia di luce in un ambiente completamente privo di luce e lo spettro di colori prodotto da un prisma. È facile misurare il flusso di luce che raggiunge l'occhio da un ambiente buio. Cambiando il flusso è possibile stimare il corrispondente cambiamento della luminosità percepita. Ebbene, si trova che la luminosità percepita cambia solo di poco in seguito a enormi variazioni del flusso. Per esempio, riducendo moltissimo il flusso si osserverà solo una

**Una macchia di luce nel vuoto, cioè una singola macchia di luce in una banda ristretta dello spettro vista in un ambiente completamente buio, possiede un colore che sembrerebbe dipendere solo dalla sua lunghezza d'onda. Il suo colore, peraltro, può venire spiegato in termini delle luminosità percepite dai tre sistemi fotorecettori dell'occhio. Misure psicofisiche dimostrano che in queste condizioni la luminosità percepita cambia molto poco per grandi variazioni del flusso, come è indicato dalla linea retta. Per esempio, se la macchia è costituita da una banda ristretta centrata, diciamo, a 600 nanometri, i tre pigmenti dei coni assorbiranno flussi molto diversi, a causa della forma delle loro curve di assorbimento. In unità arbitrarie il pigmento sensibile a grandi lunghezze d'onda assorbirà 80 unità, quello sensibile a lunghezze d'onda intermedie 20 unità e il terzo pigmento al massimo qualche decimo di unità. I valori di luminosità corrispondenti a questi flussi sono 9 per le grandi lunghezze d'onda, 8,5 per quelle intermedie e 7,5 per quelle corte. Questa combinazione di luminosità viene percepita come una luce arancio-rossastra, non percepibile in condizioni ordinarie a meno che si abbia a che fare con colori fluorescenti.**

**Quando si interpone la mano o un altro oggetto fra un proiettore di luce a grande lunghezza d'onda («rossa») e uno schermo bianco illuminato con luce bianca si osserva un'ombra di colore bluverde (azzurro). L'autore considera la descrizione delle ombre colorate, prodotte dalla luce di candela, fatta da Otto von Guericke nel 1672 come la prima osservazione importante per la comprensione del meccanismo di percezione delle immagini e dei colori. Nell'analisi a destra si assume che un proiettore illumini lo schermo con luce bianca. L'altro proiettore, equipaggiato con un filtro rosso, illumina lo schermo solo con radiazioni di grande lunghezza d'onda. Assumiamo che la luce bianca fornisca 100 unità arbitrarie di flusso a ognuno dei tre sistemi di fotorecettori. Il flusso del secondo proiettore è assorbito in proporzioni differenti dai tre sistemi di fotorecettori: 100 unità sono assorbite dal sistema sensibile alle grandi lunghezze d'onda, 50 da quello sensibile alle lunghezze intermedie e 5 dal terzo. (Una piccola quantità di radiazione di grande lunghezza d'onda compare per diffusione anche nell'ombra.) La terza colonna mostra le quantità complessive di flusso assorbite dai tre sistemi fotorecettori. Le frazioni rappresentano i rapporti ai margini dell'ombra fra il flusso proveniente da questa e il flusso proveniente dall'esterno. La quarta colonna mostra infine le luminosità percepite dai sistemi fotorecettori. La luminosità delle zone più luminose per ogni sistema si troverà vicina all'estremo più chiaro della scala di luminosità, essendo determinata dal flusso di energia raggiante allo stesso modo in cui il flusso determina la luminosità di una macchia di luce. Le triplette di luminosità all'interno dell'ombra cadono nella regione dello spazio dei colori che l'occhio percepisce come azzurro.**

piccolissima riduzione di luminosità. Se la macchia di luce è costituita da una piccola banda spettrale di grande lunghezza d'onda, intorno, diciamo, a 600 nanometri, ci si può attendere che tutti e tre i sistemi di coni assorbano della radiazione, ma quelli sensibili alle grandi lunghezze d'onda ne assorbiranno parecchia di più. Leggendo i tre valori su una scala delle luminosità percepite, le tre luminosità risultano 9 per il sistema sensibile a grandi lunghezze d'onda, 8,5 per quello sensibile a lunghezze d'onda intermedie e 7,5 per il terzo (*si veda la figura nella pagina precedente*). Questa tripletta di luminosità è percepita come un arancione rossastro, un colore raramente percepito in condizioni ordinarie, a meno che le superfici siano fluorescenti. Lo spettro, un insieme di colori straordinariamente anomalo, può essere considerato come una serie di tre gradienti continui di colore affiancati, che possiedono contemporaneamente sia le proprietà delle macchie di luce sia quelle delle superfici colorate.

A causa di queste proprietà è possibile prevedere i colori dello spettro, mentre non è possibile, come abbiamo visto, attribuire una specifica composizione spettrale, nelle condizioni abituali, di tutti i giorni, alla luce che viene riflessa da una superficie colorata.

Probabilmente la prima osservazione importante per la comprensione del meccanismo della percezione dei colori non è quella di Newton dello spettro, bensì quella delle ombre colorate, descritta nel 1672 da Otto von Guericke. «Accade - egli scrisse - che nella penombra del primo mattino si può produrre un'ombra azzurra su un pezzo di carta [tenendo] un dito o un altro oggetto ... fra una candela accesa e la carta posta sotto di essa». Ora sappiamo che questo importante esperimento descrive un esempio elementare della generazione di tre diverse luminosità su un sistema a tre recettori. Uno schema dell'esperimento con luce bianca o con luce di grande lunghezza d'onda («rossa») è descritto nella figura di questa pagina. Qui il colore dell'ombra è blu. La figura mostra che la tripletta di luminosità nell'ombra corrisponde al colore bluverde che è possibile prevedere dalla sua posizione nello spazio colore - luminosità. È ora possibile comprendere che i nostri vecchi esperimenti con le immagini in luce bianca e rossa non sono che un procedimento per ottenere colori tipo ombra colorata in molte sfumature e nelle immagini anziché nelle ombre. I colori visti in una proiezione in bianco e rosso possono venire facilmente previsti estendendo l'analisi fatta per giustificare il colore dell'ombra di von Guericke. Per dimostrare questo punto, a pagina 68, sono riprodotte le immagini di selezione con filtri rosso e verde usate per una proiezione a colori in luce rossa e bianca (in questa dimostrazione non si è fatto alcun tentativo per riprodurre i colori effettivamente visti nella proiezione). La proiezione in bianco e rosso è stata fotografata attraverso una combinazione filtro-pellicola retinex sensibile a bande di grande, media e corta lunghezza d'onda. Le tre immagini sono riprodotte sotto la coppia di selezioni cromatiche (a grande e media lunghezza d'onda) usate per ottenere la proiezione in luce rossa e bianca. Il punto significativo è che quando l'occhio, utilizzando il proprio sistema retinex, vede sullo schermo l'immagine rossa e bianca ottiene per ogni punto dell'immagine una tripletta di luminosità che assomiglia a quella che otterrebbe guardando direttamente il soggetto usato per le riprese. In questo importante punto d'incontro fra le ombre azzurre e le immagini colorate il lavoro estensivo di determinazione e moltiplicazione dei rap-

SELEZIONE A GRANDI LUNGHEZZE D'ONDA

SELEZIONE A MEDIE LUNGHEZZE D'ONDA

FILTRO ROSSO

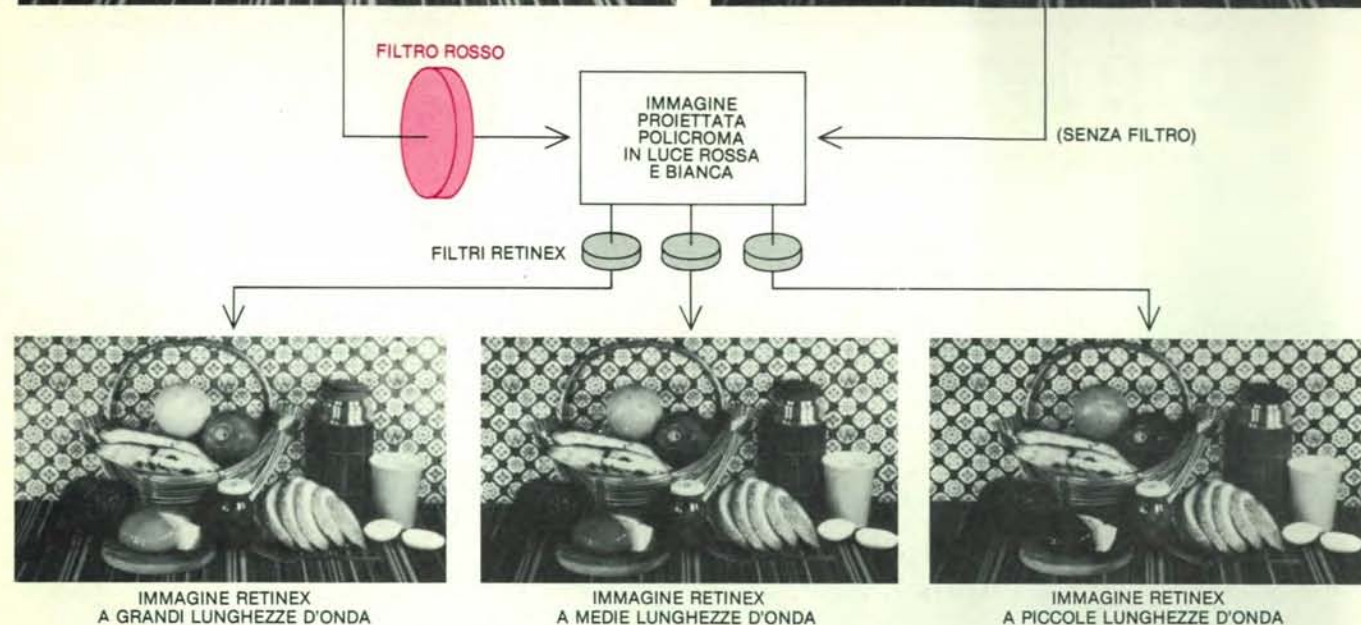
IMMAGINE PROIETTATA POLICROMA IN LUCE ROSSA E BIANCA

(SENZA FILTRO)

FILTRI RETINEX

IMMAGINE RETINEX A GRANDI LUNGHEZZE D'ONDA

IMMAGINE RETINEX A MEDIE LUNGHEZZE D'ONDA

IMMAGINE RETINEX A PICCOLE LUNGHEZZE D'ONDA

**Immagini retinex di proiezioni in luce rossa e bianca mostrano che queste producono triplette di luminosità in accordo con le sensazioni di colore provate durante la proiezione. Le due fotografie in alto sono riproduzioni delle selezioni nelle bande a grande (*a sinistra*) e media (*a destra*) lunghezza d'onda di una natura morta. La prima immagine è stata proiettata su uno schermo attraverso un filtro a banda passante di grande lunghezza d'onda (rosso). La seconda è stata proiettata in sovrapposizione alla prima con la luce prodotta da una lampada a filamento di tungsteno. L'immagine sullo schermo è stata poi fotografata con tre combinazioni filtro-pellicola riproducenti le sensibilità spettrali dei sistemi fotorecettori. Queste «immagini retinex» sono riprodotte in basso: a sinistra quella relativa alle grandi lunghezze d'onda, al centro quella relativa alle lunghezze intermedie e a destra quella relativa alle piccole lunghezze d'onda. I colori visti nella proiezione in luce bianca e rossa sono quelli che ci si può attendere dalle loro triplette di luminosità. La mela risulta chiara nell'immagine ripresa a grande lunghezza d'onda e più scura nelle riprese nelle altre bande dello spettro. L'arancia è gradatamente meno chiara nelle tre riprese. È notevole che con i propri sistemi retinex l'osservatore possa vedere una tazza blu, un cestino marrone e delle banane gialle con differenze di luminosità tanto piccole da costituire un problema per la loro resa con i procedimenti di fotoincisione usati comunemente.**

porti di riflettanza determina la luminosità di ogni zona. In conclusione, tutti questi principi vengono applicati nella visione ternaria di ogni giorno, nel corso della quale vengono formate tre distinte immagini per i tre diversi sistemi sensibili, immagini che vengono confrontate per produrre la sensazione del colore.

Questo programma di ricerche teoriche e verifiche sperimentali ha avuto inizio 25 anni fa, in seguito all'osservazione che le energie dei proiettori in luce rossa e bianca potevano venire cambiate senza che cambiassero i colori dell'immagine proiettata. Questa osservazione confutava la spiegazione semplicistica della visione dei colori in termini di contrasto, affaticamento e influenza degli oggetti circostanti, portando a formulare l'idea fondamentale di due sistemi indipendenti di formazione dell'immagine sensibili a grandi e corte lunghezze d'onda, idea successivamente evolutasi in quella di tre sistemi retinex indipendenti, dimostrati dalle esperienze con i collage. L'idea che la variabile determinante fosse la percentuale di radiazione riflessa in ognuna delle tre bande dello spettro e la tecnica della sua misurazione hanno portato ad ammettere che le luminosità mantengono una graduatoria indipendente nelle bande a grande e piccola lunghezza d'onda. La tecnica sperimentale si è poi evoluta dalla proiezione di un'immagine in bianco e nero alla disposizione in collage di pezzi di carta colorata. L'evidenza della stabilità e della costanza delle luminosità di tutti i pezzi del collage quando viene illuminato con una singola banda dello spettro di intensità variabile sottolinea il fatto che ogni pezzo di carta è caratterizzato da tre riflettanze in tre diverse bande dello spettro e che tali riflettanze sono in rapporto con una grandezza di preciso interesse biologico: la luminosità.

I colori possono venire disposti nello spazio delle luminosità, definito da tre assi di luminosità in bande di radiazione a grande, media e piccola lunghezza d'onda. Tutti i colori visibili si dispongono in questo spazio, indipendentemente dal flusso energetico, e corrispondono a una terna di coordinate di luminosità. È necessario tener presente che la realtà dei colori va ricercata in questo spazio. Quando il collage a colori viene fotografato o misurato con un illuminazione non uniforme la riflettanza nelle fotografie non corrisponde al colore, ma la luminosità continua a corrispondervi. I tre insiemi di rapporti fra integrali ai margini e i prodotti di tali integrali entro l'insieme di superfici risultano essere i determinanti fisici delle relazioni che si instaurano fra il sistema biologico e il mondo esterno.

# La struttura del mezzo interstellare

*Dall'analisi fotografica dei segnali radio emessi dal gas interstellare sembra che esso non formi nubi, come generalmente si credeva, ma sia disposto ordinatamente in enormi gusci*

di Carl Heiles

Mezzo secolo fa le osservazioni spettroscopiche di stelle lontane rivelarono che la loro luce è arrossata dal passaggio attraverso il gas e la polvere che permeano lo spazio interstellare. La scoperta di questa materia interstellare sollevò vari problemi, ancora per lo più insoluti. Uno dei più importanti riguardava la struttura del mezzo interstellare. In qualche regione del cielo vicino a stelle molto calde gli atomi del gas interstellare sono ionizzati, privati cioè di uno e più elettroni, ed emettono una luce caratteristica. I primi quattro decenni di osservazioni di queste regioni ionizzate suggerirono che la materia interstellare non è distribuita uniformemente nello spazio, ma si trova ammassata in grandi nubi grossolanamente sferiche sparse casualmente nella Galassia. Queste osservazioni diedero origine a un modello teorico del mezzo interstellare noto come il «modello a nubi». Secondo questo modello le nubi di gas interstellare avrebbero un diametro di circa 14 parsec e disterebbero tra loro circa 30 parsec. (Un parsec corrisponde a 3,26 anni luce.)

La maggior parte della materia interstellare è però di gran lunga troppo fredda per emettere energia alle lunghezze d'onda della luce visibile e, pertanto, rimane inosservabile ai telescopi ottici. Nell'ultimo decennio circa gli astronomi hanno potuto estendere le loro osservazioni del mezzo interstellare ad altre regioni dello spettro, comprendendo i raggi X, l'ultravioletto, l'infrarosso e le onde radio. Queste osservazioni, e in particolare le mappe dettagliate della distribuzione spaziale dell'idrogeno atomico ottenute alla lunghezza d'onda radio di 21 centimetri, hanno cambiato radicalmente le nostre concezioni sulla natura e la struttura del gas tra le stelle.

Un atomo di idrogeno è costituito da un protone e da un elettrone, rotanti entrambi attorno al proprio asse. Gli spin possono essere paralleli tra loro o antiparalleli. Se sono paralleli, l'elettrone può invertire il proprio spin così da diventare antiparallelo a quello del protone. Questa transizione è accompagnata da un emissione di radiazione alla frequenza di 1420,4057 megahertz, che corrispondono a una lunghezza d'onda radio di 21 centimetri circa. L'idrogeno, che è l'elemento più abbondante dell'universo, costituisce il 70 per cento circa della massa del gas interstellare. Più della metà dell'idrogeno è in forma molecolare ($H_2$) oppure è ionizzato, cioè il protone è separato dall'elettrone. L'idrogeno molecolare e l'idrogeno ionizzato si trovano però solo in una piccola frazione del volume dello spazio interstellare, dove il mezzo è relativamente denso. L'idrogeno che permea il resto del volume è in forma atomica.

Le osservazioni dell'idrogeno atomico a lunghezza d'onda radio presentano un vantaggio decisivo sulle osservazioni nel visibile. Infatti nell'ottico si studia l'idrogeno interstellare osservando come il gas assorbe la luce stellare; pertanto le osservazioni del gas a quelle lunghezze d'onda sono limitate alle regioni del cielo dove c'è una stella adatta dietro al gas. Nel radio si studia l'idrogeno interstellare osservando la radiazione emessa dal gas stesso, e non è necessaria una sorgente di radiazione di fondo. Pertanto queste osservazioni possono essere fatte in qualsiasi regione e possono rivelare la distribuzione del gas in tutto il cielo.

*I primi rilevamenti celesti*

I primi tentativi di ottenere una mappa della distribuzione dell'idrogeno si limitarono a osservazioni del gas nel piano centrale della Galassia. Presto però fu chiaro che vi era così tanto gas nel piano galattico che non si poteva distinguere il gas di una qualsiasi struttura da tutto il resto del gas lungo la linea di vista. I rilevamenti successivi si concentrarono maggiormente su regioni non appartenenti al piano galattico, studiando aree di cielo di ampiezza compresa tra 10 e 20 gradi. Secondo il modello a nubi del mezzo interstellare dovrebbero essere visibili dozzine e forse centinaia di nubi di gas in un'area di cielo di diametro angolare compreso tra 10 e 20 gradi. I primi rilevamenti del mezzo interstellare fuori dal piano galattico mostrarono però un fatto sorprendente: le strutture gassose osservate erano così ampie su scala angolare che si estendevano oltre i limiti delle regioni considerate. Nonostante apparissero anche strutture su scala più ridotta, queste sembravano contenute in quelle più grandi. Era chiaro che non si sarebbero potuti capire gli aspetti strutturali minori finché non si fossero compresi quelli più grandi.

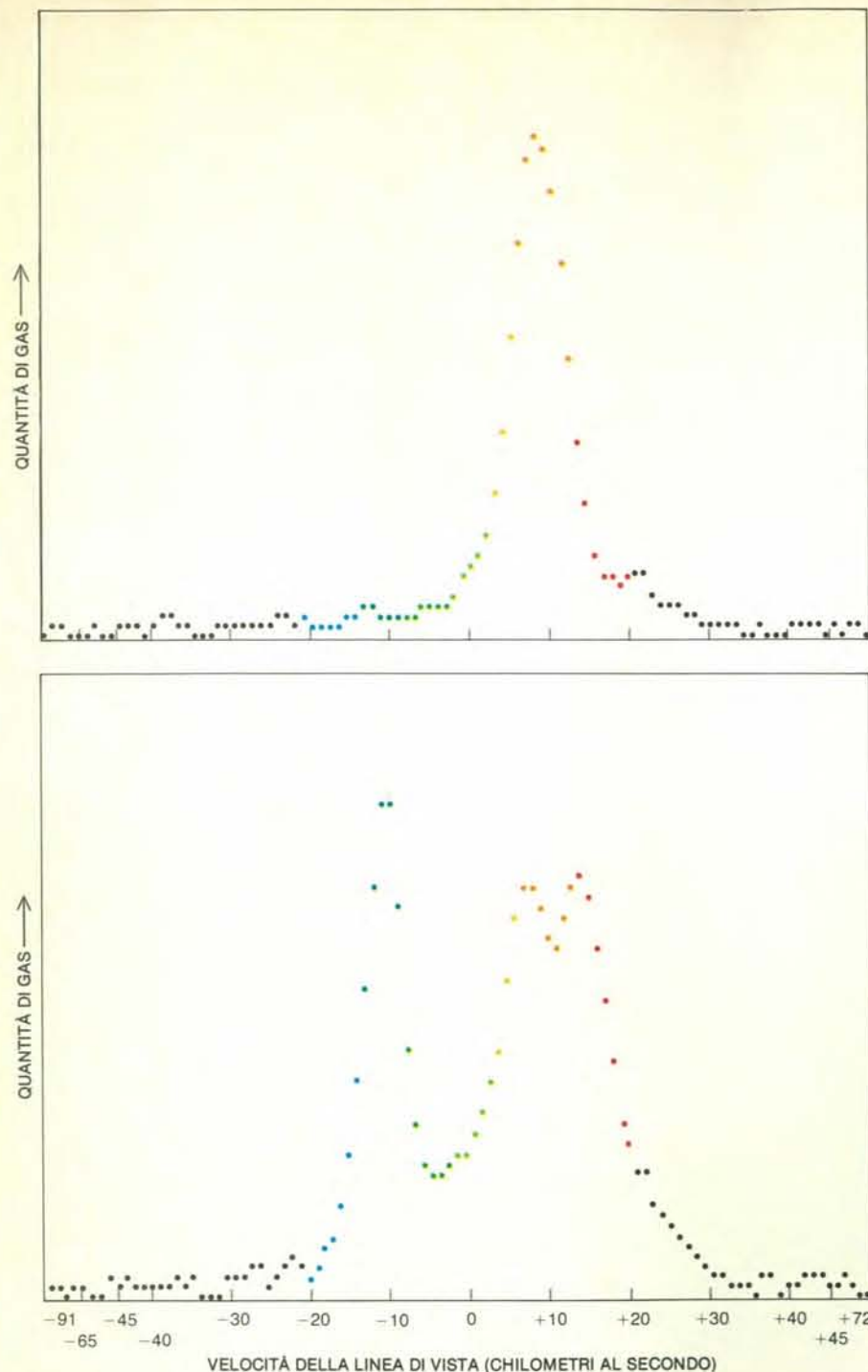
Dato che non si conosceva l'estensione angolare totale delle strutture di grandi dimensioni, si dovevano evidentemente costruire mappe di aree di cielo più estese. Inoltre, l'astronomia è una disciplina statistica per sua stessa natura: infatti gli astronomi non possono modificare le condizioni nello spazio in modo da condurre esperimenti sul comportamento dei sistemi celesti, così che devono accontentarsi di osservare gli oggetti nella varietà di condizioni che si offrono naturalmente. Per osservare un campione statisticamente significativo delle grandi strutture del mezzo interstellare, si dovrebbero costruire mappe di aree realmente molto estese. I miei colleghi dell'Università della California a Berkeley e io decidemmo di fare un rilevamento di tutto il cielo o, almeno, di tutto il cielo osservabile dall'osservatorio radioastronomico universitario di Hat Creek, circa 280 chilometri a nord di Sacramento presso il Lassen Volcanic National Park.

Con un radiotelescopio di 26 metri di diametro operammo un rilevamento del cielo visibile da Hat Creek, escludendo una fascia di 20 gradi di ampiezza centrata sul piano galattico, dove vi è tanto gas che è impossibile distinguere singole strutture. Per coprire una regione di cie-

**Questa sintesi fotografica del mezzo interstellare mostra come il gas del mezzo si insinua nello spazio interstellare in lunghi filamenti. L'illustrazione è stata sintetizzata artificialmente dai dati radio a 21 centimetri ottenuti con il telescopio da 26 metri dello Hat Creek Radio Observatory dell'Università della California. L'autore e i suoi colleghi fecero un rilevamento del mezzo interstellare su tutto il cielo visibile dalla California escludendo un'ampia striscia centrata sul piano galattico dove vi è troppo gas per distinguere le singole strutture. La grande parabola nera a sinistra è la parte del cielo meridionale invisibile dalla latitutine dell'osservatorio; il piccolo rettangolo nero presso 120 gradi di longitudine galattica e 30 gradi di latitudine galattica è il polo celeste nord, dove non si può puntare il telescopio a causa delle sue caratteristiche costruttive. Si noti che una parte di longitudine galattica è rappresentata due volte, così che si può vedere la continuità degli aspetti strutturali ai bordi della mappa. La brillantezza di ogni filamento nell'illustrazione rappresenta la quantità di gas presente in quella regione. Il colore di ogni dettaglio rappresenta la velocità del gas e la direzione in cui si sta muovendo. C'è una lenta progressione di colore attraverso lo spettro al variare della velocità: con il rosso si rappresenta gas che si allontana dal sistema solare a 20 chilometri al secondo, in giallo-verde si indica gas senza componenti di velocità lungo la linea di vista, mentre con il blu si rappresenta gas che si avvicina al sistema solare alla velocità di 20 chilometri al secondo. I colori intermedi indicano velocità intermedie. Si può vedere come presso il piano della Galassia il colore vada variando in generale dal blu al rosso con la longitudine galattica. Questo alternarsi dei colori riflette il fatto che la Galassia non ruota come ruoterebbe un disco rigido.**

QUANTITÀ DI GAS ⟶

QUANTITÀ DI GAS ⟶

VELOCITÀ DELLA LINEA DI VISTA (CHILOMETRI AL SECONDO)

**Con il radiotelescopio di Hat Creek furono misurati profili di velocità del mezzo interstellare in 140 000 posizioni nel cielo. Ogni profilo rappresenta la quantità di gas che si muove lungo la linea di vista dell'osservatore a velocità diverse. Tutti i profili furono combinati per ottenere le sintesi fotografiche. Velocità positive indicano che il gas si sta allontanando dal sistema solare, velocità negative indicano gas in avvicinamento. Ogni puntino rappresenta una misura della quantità di gas in ognuno dei 100 piccoli intervalli di velocità. Dato che nel mezzo interstellare solo una quantità di gas molto ridotta si muove con velocità molto elevate rispetto al sistema solare, l'ampiezza dell'intervallo di velocità è maggiore per velocità maggiori di più o meno 45 chilometri al secondo. La forma del profilo di velocità del gas interstellare è la stessa in ogni posizione nel cielo. Costruendo la sintesi fotografica a colori si è immaginato che la lunghezza d'onda della luce visibile crescesse da sinistra verso destra attraverso il profilo radio di velocità, facendo corrispondere il blu puro alla velocità di meno 20 chilometri al secondo e il rosso puro alla velocità di più 20 chilometri al secondo. Il profilo di velocità in alto fu misurato per una piccola regione presso il punto di 181,7 gradi di longitudine galattica e —30,4 gradi di latitudine galattica. In quella posizione la maggior parte del gas si allontana dal sistema solare quasi con la stessa velocità. Il profilo di velocità è stretto ed è rappresentato da un colore relativamente puro; nella sintesi fotografica delle pagine precedenti la regione è arancio-rossastra. Il profilo di velocità in basso fu misurato per una piccola regione presso il punto di 205,7 gradi di longitudine galattica e —22,6 gradi di latitudine galattica. In tale posizione una parte del gas si avvicina al sistema solare (*picco stretto a sinistra*) e una parte poco più grande se ne allontana (*picchi a destra*). Nella sintesi fotografica delle pagine precedenti la regione è rosso-biancastra.**

lo così vasta fu necessario puntare il telescopio verso 140 000 posizioni distinte circa. La sensibilità delle osservazioni cresce con il tempo dedicato a ogni posizione; abbiamo dovuto dedicare circa due minuti a ogni posizione per poter ottenere dati con la sensibilità necessaria. Le osservazioni richiesero un totale di sette mesi di tempo effettivo di telescopio, distribuito su un arco di due anni.

### *Una ricchezza di nuovi dati*

In ogni posizione raccogliemmo informazioni sia sull'ammontare complessivo del gas sia sul suo stato di movimento. Nonostante che ogni atomo di idrogeno emetta energia alla frequenza di 1420,4057 megahertz, la frequenza a cui il telescopio riceve la radiazione è spostata per effetto Doppler dovuto al moto del gas rispetto all'osservatore. L'ammontare dello spostamento di frequenza dipende solo dalla direzione di movimento e dalla velocità del gas lungo la linea di vista: il gas che si allontana dall'osservatore appare a frequenza minore, quello che si avvicina appare a frequenza maggiore. Per convenzione si attribuiscono velocità positive a sorgenti che si allontanano dal sistema solare e velocità negative a sorgenti che si avvicinano.

Abbiamo misurato l'intensità della radiazione proveniente dal gas a 100 frequenze diverse, corrispondenti cioè a 100 intervalli di velocità diversi. La quantità di gas che si muove con una data velocità è proporzionale all'intensità della radiazione proveniente dal gas che si muove con tale velocità. In molti casi c'è più di una struttura individuale di gas lungo la linea di vista, con le varie strutture localizzate a distanze diverse dal sistema solare. Normalmente ogni struttura avrà uno spostamento Doppler diverso da quello delle altre, e pertanto apparirà in un intervallo di velocità diverso.

Qualsiasi rilevamento esteso a tutto il cielo fornisce grandi quantità di dati, che costituiscono di per sè un problema. Nel nostro caso il problema era particolarmente grave. Dato che misuravamo lo ammontare del gas in 100 intervalli di velocità per ciascuna delle 140 000 posizioni nel cielo, ottenevamo un totale di 14 milioni di punti, ciascuno dei quali è localizzato in uno spazio tridimensionale di longitudine galattica, latitudine galattica e velocità. È un vero problema organizzare tanti dati in modo che siano rapidamente interpretabili dalla mente umana. Normalmente in radioastronomia si presentano i dati sotto forma di una mappa in cui la quantità di gas presente è rappresentata con curve di densità costante, proprio nello stesso modo con cui si rappresentano sulle carte topografiche le altezze delle montagne mediante curve di quota costante. Dapprima compilammo una mappa di questo genere per i nostri dati, ma le strutture a grande scala che volevamo studiare erano nascoste dal labirinto delle curve di livello delle strutture minori interne. Occorreva un modo migliore per presentare i dati.

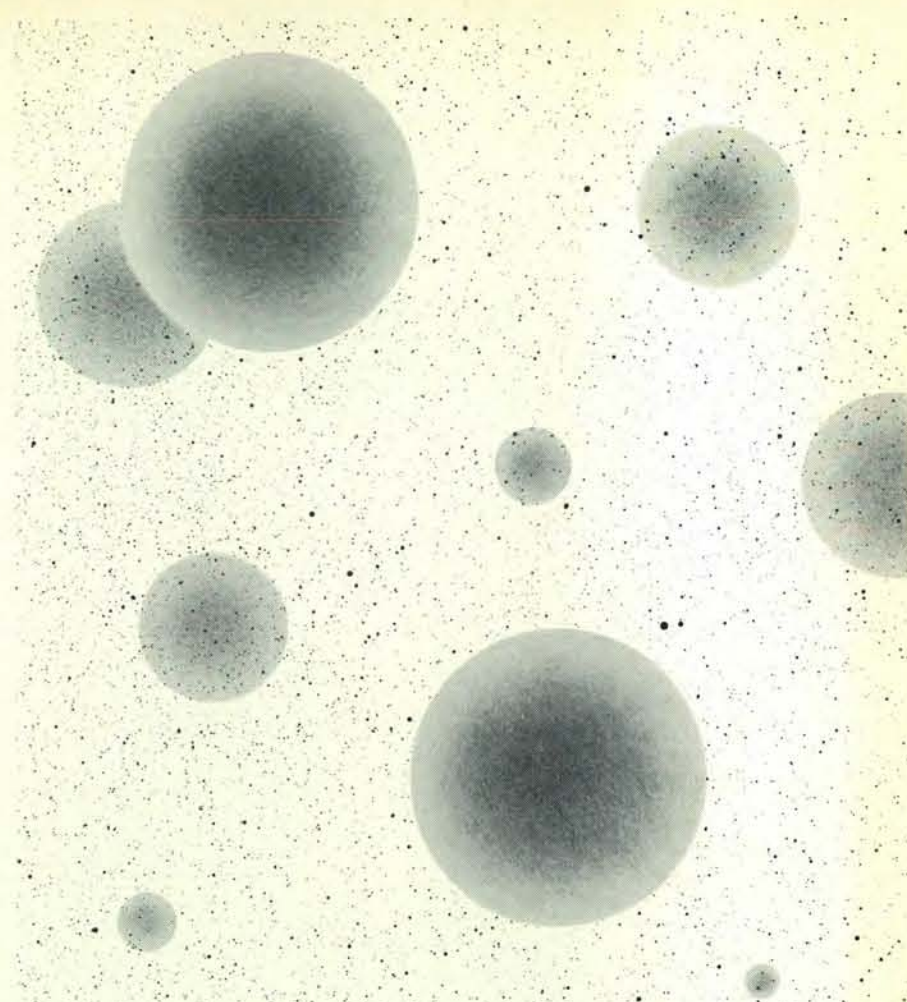


Gli astronomi che osservano il cielo nel visibile raccolgono regolarmente dati relativi a oggetti di grandi dimensioni contenenti strutture di piccola scala registrandoli su una pellicola fotografica. I miei colleghi e io avevamo bisogno di una macchina capace di sintetizzare una «fotografia» dai dati numerici che avevamo raccolto e immagazzinato su nastro magnetico. Fortunatamente una macchina di tal genere era stata messa a punto per altri scopi al Dipartimento di astronomia della Princeton University. Inoltre, numerosi astronomi di Princeton lavorano anch'essi sulla struttura del mezzo interstellare e quindi la collaborazione è nata spontaneamente.

### *La sintetizzazione di radiofotografie*

Lavorando con Edward B. Jenkins di Princeton, trascurammo dapprima l'informazione contenuta nei singoli intervalli di velocità e realizzammo semplicemente una sintesi fotografica della quantità totale di gas compresa nell'ampio intervallo di velocità tra meno 20 chilometri al secondo e più 20 chilometri al secondo. Quasi tutto il gas è contenuto in questo grande intervallo (*si veda l'illustrazione nelle due pagine successive*). La prima cosa che notammo sulla fotografia così sintetizzata fu la tendenza del gas a organizzarsi non in singole nubi, ma in lunghe strutture filamentari. Molti filamenti hanno grandi dimensioni angolari; alcuni si estendono per un quarto di cielo. Non si può determinare però la lunghezza reale in parsec senza conoscerne la distanza dal sistema solare.

Le osservazioni radio del gas non forniscono di per sè dati sulla distanza dei filamenti dal sistema solare. Per ottenere questa informazione è necessario connettere le osservazioni del gas a quelle di stelle di cui si può determinare indipendentemente la distanza. Il modo più immediato e diretto per correlare i due tipi di osservazioni consiste nello studiare le righe di assorbimento prodotte nello spettro di una stella dal gas interstellare che si trova tra la stella e il sistema solare. Le righe di assorbimento sono il risultato del fatto che il gas interstellare a bassa temperatura assorbe la radiazione proveniente dalla stella più calda dietro di sè alle stesse lunghezze d'onda a cui emetterebbe il gas stesso, se riscaldato a una temperatura sufficientemente elevata.

Jenkins e Blair D. Savage dell'Università del Wisconsin hanno studiato, negli spettri di stelle, le righe di assorbimento prodotte dall'idrogeno atomico interstellare. Poiché sia le righe di assorbimento sovrapposte allo spettro della stella, sia le righe di emissione nel radio a 21 centimetri sono prodotte da idrogeno atomico, un confronto tra l'intensità delle righe di assorbimento e di emissione rivelerà se tutto l'idrogeno è davanti alla stella o se parte di esso si trova dietro alla stella, ove non avrebbe nessun effetto sullo spettro stellare. Determinata poi la distanza della stella con le usuali tecniche astronomiche, si può essere sicuri che il gas davanti alla stella non dista dal sistema solare più del valore trovato per la stella stessa.

**Il modello a nubi della struttura del gas interstellare fu proposto sulla base delle osservazioni nell'ottico. Il modello proponeva che il gas si trovasse ammassato in nubi singole (*in grigio*) di sette parsec di raggio distribuite nello spazio con una distanza media tra loro di 30 parsec. (Un parsec corrisponde a 3,26 anni luce.) La distanza media tra le stelle (*punti neri*) è di 0,4 parsec.**

Le righe di assorbimento dell'idrogeno atomico interstellare si trovano tutte nella regione ultravioletta dello spettro, che non è osservabile dalla terra. Sul satellite *Copernicus* in orbita intorno alla Terra si trova però un telescopio adatto per osservare a quelle lunghezze d'onda, anche se le sue possibilità osservative sono limitate alle stelle ad alta temperatura che sono molto brillanti nell'ultravioletto. Stelle di questo tipo sono rare, ma quelle esistenti forniscono informazioni fondamentali non ottenibili altrimenti.

Una vasta area di cielo contenente stelle calde si estende per quasi 30 gradi attraverso la costellazione di Orione, comprendendo anche la Grande Nebulosa in Orione. Tale regione si trova presso 210 gradi di longitudine galattica e —20 gradi di latitudine galattica. Tutte le stelle in quella zona distano circa 500 parsec dal sistema solare. Gli spettri ultravioletti delle stelle mostrano che in questa regione c'è poco gas davanti alle stelle. Le osservazioni a lunghezze d'onda radio mostrano però che nella stessa regione c'è abbondanza di gas interstellare e che tutto il gas fa parte di una sola grande struttura. Queste osservazioni ci rivelano la presenza di una grande area di cielo in cui vi è poco gas a meno di 500 parsec dal sistema solare. Le osservazioni contraddicono le previsioni ottenute seguendo il modello a nubi, secondo il quale la distanza media tra le nubi interstellari sarebbe di 30 parsec.

### *Il campo magnetico galattico*

Un modo meno diretto per associare osservazioni di stelle con osservazioni del mezzo interstellare fu scoperto casualmente realizzando la prima sintesi fotografica. Alcuni filamenti del gas apparivano allineati con le linee di forza del campo magnetico della Galassia. Il campo magnetico allinea grani allungati di polvere interstellare, che sono associati al gas, così che la polvere agisce come un filtro polarizzante sulla luce stellare che l'attraversa. Si ritiene che la direzione di polarizzazione sia parallela alle linee di forza del campo magnetico; pertanto una misura della direzione di polarizzazione

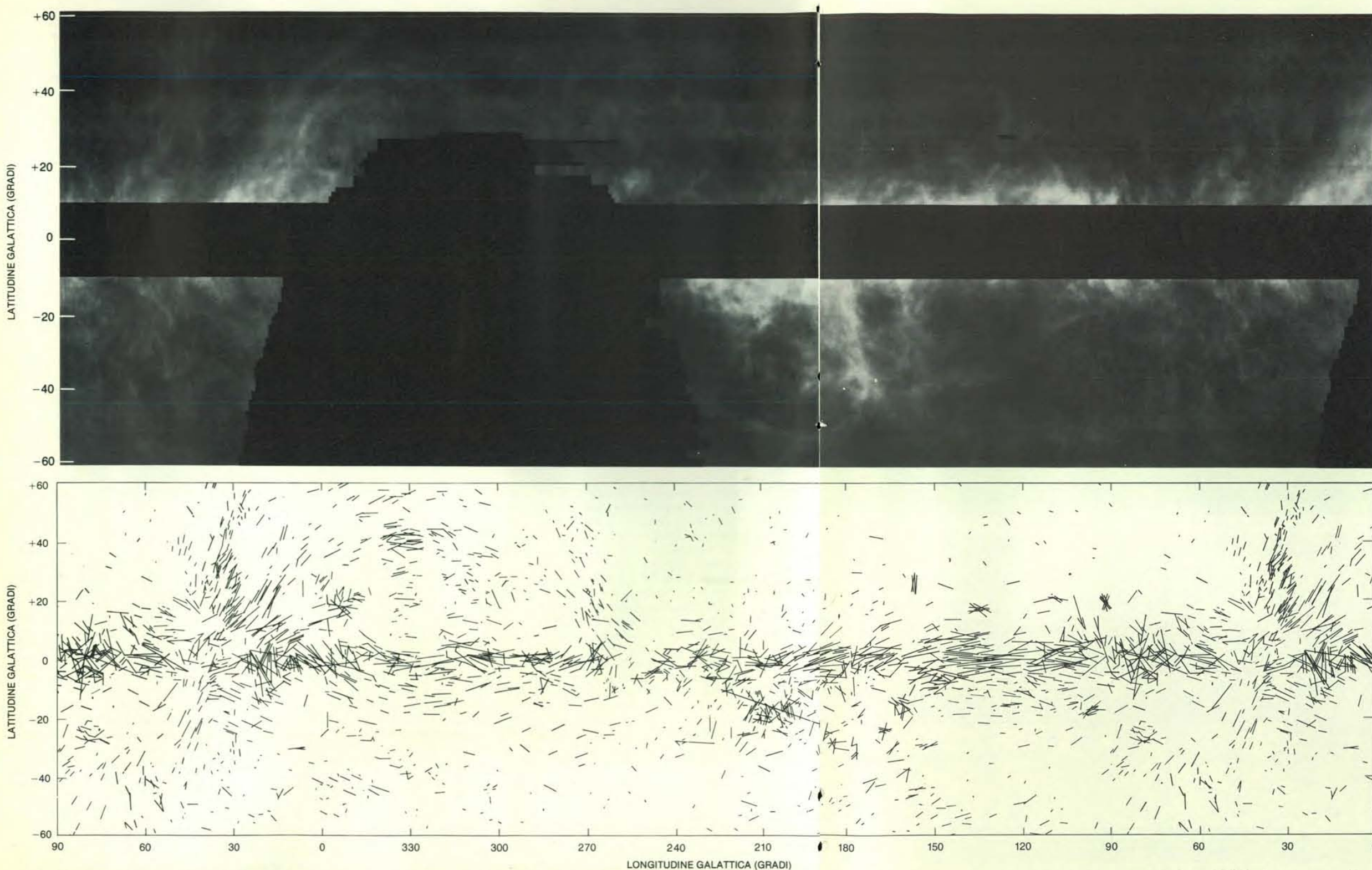
della luce stellare in effetti è anche una misura della direzione del campo magnetico galattico, mediata sulle singole concentrazioni di materia interstellare presenti tra il sistema solare e la stella che emette la luce. Se c'è una sola concentrazione di materia interstellare tra il sistema solare e la stella, la direzione di polarizzazione della luce stellare indica senza ambiguità la direzione del campo magnetico. La polarizzazione della luce stellare è stata misurata per un gran numero di stelle sparse in tutto il cielo. Le misure sono state raccolte e organizzate in una mappa da D.S. Mathewson e K.L. Ford degli osservatori di Mount Stromlo e Siding Spring dell'Università nazionale australiana (*si veda l'illustrazione in basso in queste due pagine*).

Nonostante il gran numero di stelle rappresentato sulla mappa di polarizzazione, rimangono ancora regioni di cielo non adeguatamente coperte. Vi sono anche molte regioni in cui la polarizzazione sembra essere caotica, probabilmente per la presenza di più di una concentrazione di materia interstellare lungo la linea di vista. Questi problemi fanno sì che sia difficile proporre regole generali sulla correlazione esatta tra gas interstellare e campo magnetico galattico. Vi è una regione, però, a latitudine galattica positiva presso 30 gradi di longitudine galattica, in cui l'allineamento del gas con il campo magnetico è evidente. Credo che si possa concludere da questa chiara relazione che i filamenti di gas sono effettivamente allineati con il campo magnetico galattico.

Le distanze delle stelle in questa regione indicano che il gas dista dal sistema solare circa 100 parsec. I lunghi filamenti di gas si estendono ad arco nel cielo per 50 gradi circa, pertanto devono essere lunghi circa 100 parsec. Hanno una ampiezza di pochi gradi, corrispondente grossolanamente a 14 parsec, cioè al diametro di una «nube standard» secondo il modello a nubi del mezzo interstellare basato sulle vecchie osservazioni ottiche. Sembra molto improbabile che l'accordo tra l'ampiezza dei filamenti e la previsione del modello a nubi sia casuale. Dato che le osservazioni ottiche del gas interstellare dipendevano dalla possibilità di vedere il gas illuminato dalle stelle di fondo e dato il campionamento approssimativo del gas interstellare ottenibile con questa tecnica, è improbabile che si vedano due o più stelle allineate lungo un filamento. A seguito di ciò le osservazioni ottiche del mezzo interstellare fornivano una misura delle strutture gassose equivalente alla larghezza di un filamento piuttosto che alla sua lunghezza.

*Intrecci di filamenti*

La prima sintesi fotografica della quantità totale del gas nello spazio interstellare diede dunque notevoli informazioni sulla struttura del mezzo interstellare. I miei colleghi e io pensammo che sarebbe stato possibile dedurre altre informazioni sul gas esaminando il moto dei suoi filamenti. Decidemmo di rappresentare velocità diverse del gas con colori differenti. Dato che nel visibile la radiazione emessa da un oggetto che si allontana dal sistema solare è spostata per effetto Doppler verso la regione rossa dello spettro, decidemmo di rappresentare in rosso il gas in allontanamento dal sistema solare, in giallo-verde il gas che non presenta spostamenti Doppler significativi e in blu il gas in avvicinamento al sistema solare. Rappresentammo tutto il gas dotato di velocità comprese nell'ampio intervallo tra meno 20 chilometri al secondo e più 20 chilometri al secondo; l'intensità di ogni colore indicava la quantità di gas in movimento con velocità nell'intervallo corrispondente.

La caratteristica a grande scala più evidente in questa sintesi fotografica a colori è il fatto che a basse latitudini galattiche il colore del gas si alterna dal rosso al blu al crescere della longitudine galattica. Questo alternarsi di colori è dovuto alla rotazione differenziale della Galassia. La Galassia, infatti, non ruota rigidamente come un disco sul fonografo; se lo facesse non ci sarebbero movimenti relativi di una parte della Galassia rispetto a un'altra. Invece le regioni interne della Galassia ruotano più velocemente di quelle esterne. Questa rotazione differenziale implica che le stelle lontane tra loro hanno un moto relativo maggiore di quello di stelle tra loro vicine. Le velocità differenziali a grande scala delle stelle rispetto al Sole possono essere messe in relazione con la velocità media del gas interstellare così da fornire una rozza stima della distanza del gas dal sistema

**I filamenti sono allineati con il campo magnetico della Galassia, come si vede confrontando una sintesi fotografica del mezzo interstellare (*in alto*) e una mappa della polarizzazione della luce stellare (*in basso*). La polvere che si trova tra una stella e il sistema solare polarizza la luce emessa dalle stelle di fondo. Ogni linea rappresenta una misura del grado di polarizzazione della luce stellare; la lunghezza della linea è direttamente proporzionale al grado di polarizzazione, e la direzione della linea indica quella della polarizzazione. Il campo magnetico galattico è parallelo alla direzione di polarizzazione. Archi ben definiti nel campo magnetico a latitudini galattiche positive presso 30 gradi di longitudine galattica sono ben allineati con filamenti di gas vicini. Zone in cui il campo è caotico e i filamenti non sono ben allineati possono essere regioni con numerose concentrazioni di gas tra stella e sistema solare.**

solare. Le stime della distanza del gas così ottenute concordano con le stime indipendenti delle distanze determinate con le altre tecniche descritte.

### *L'individuazione di singoli filamenti*

Su scala più ridotta la sintesi fotografica a colori rivelava che il gas presenta una struttura di velocità molto dettagliata. In generale un singolo filamento non è caratterizzato da una sola velocità. In effetti, non ci sembrava che la velocità variasse con la lunghezza e la larghezza del filamento in modo sistematico. Parti del filamento sembravano torcersi e avvolgersi così che la sua struttura era caratterizzata da variazioni di velocità apparentemente casuali. Tali variazioni fornivano dati, sui moti all'interno del filamento, molto più ricchi di quelli precedentemente disponibili, ma il tipo di informazione in essi contenuta lasciava insoddisfatto chi cercava di comprendere la struttura complessiva del gas interstellare; è difficile arrivare a una interpretazione o a una spiegazione diretta di un miscuglio di numeri a caso.

Nel tentativo di dare un senso alla struttura intricata del mezzo interstellare decidemmo di esaminare il gas in intervalli di velocità più piccoli. Dividemmo l'intervallo complessivo di velocità da meno 20 chilometri al secondo a più 20 chilometri al secondo in 20 intervalli di velocità di due chilometri al secondo di ampiezza ciascuno e sintetizzammo una fotografia per ognuno di essi. Visto che la figura a colori comprendente l'intervallo completo ci aveva mostrato variazioni di velocità apparentemente casuali su scala angolare relativamente ridotta, non ci aspettavamo risultati particolarmente significativi dalla nostra serie di fotografie a stretto intervallo di velocità. Tali previsioni erano errate. Questo passo ulteriore fu probabilmente la chiave che ci permise di capire la vera natura della struttura del gas interstellare e il tipo di fenomeni fisici che modellarono tale struttura; ci consentì di vedere singole strutture di gas senza interferenze dovute ad altro gas lungo la linea di vista.

Non è possibile mostrare qui tutte le 20 sintesi fotografiche, perciò ho scelto solo le due che mettono in luce un aspetto interessante capace di illustrare bene le mie argomentazioni. Le due sintesi fotografiche appaiono nella pagina a fronte. Quella in alto copre l'intervallo tra più 16,4 e 18,4 chilometri al secondo; quella in basso si riferisce all'intervallo da più 3,7 a più 5,7 chilometri al secondo. Entrambe mostrano una grande struttura a forma ellittica di filamenti gassosi che si estende per circa 50 gradi ed è centrata in corrispondenza a 200 gradi di longitudine galattica e —40 gradi di latitudine galattica. Confrontando le due sintesi fotografiche si osserva che tale struttura presenta un grande diametro a velocità minori e un piccolo diametro a velocità maggiori. Restammo sorpresi trovando una struttura tanto coerente estesa a una regione angolare di cielo così ampia per un intervallo di velocità così grande. Ancora più sorprendente è il fatto che di questa struttura si possa dare una semplice spiegazione.

Consideriamo un guscio sferico di gas che si espande radialmente dal suo centro verso l'esterno. Un osservatore esterno al guscio vedrà uno spostamento Doppler del gas in movimento che dipende dalla proiezione del moto del gas lungo la linea di vista. Il gas sul bordo più vicino del guscio e quello sul bordo più lontano (allineati sulla linea di vista passante per il centro) saranno rispettivamente in avvicinamento diretto e in allontanamento diretto relativamente all'osservatore e presenteranno pertanto il maggiore spostamento Doppler. Il gas all'estremità del guscio alla massima distanza angolare del centro si muoverà quasi ad angolo retto rispetto alla linea di vista e avrà pertanto lo spostamento Doppler minore. Il gas che si trova tra il centro del guscio e la sua estremità si muoverà a velocità intermedie lungo la linea di vista e presenterà spostamenti Doppler intermedi. A seguito di ciò l'osservatore dovrebbe vedere in ogni piccolo intervallo di velocità un anello di gas le cui dimensioni angolari diminuiscono al crescere della componente della velocità lungo la linea di vista, e questo è proprio ciò che si osserva. La struttura filamentare ellittica che vediamo ha il suo diametro minimo in corrispondenza della velocità massima e presenta il suo diametro massimo quando la velocità è quasi nulla. Il guscio si sta espandendo con una velocità di 23 chilometri al secondo.

**Un anello ellittico di gas è riconoscibile presso il centro in basso in entrambe queste sintesi fotografiche, riprodotte qui come stampe in negativo. L'illustrazione in alto è una sintesi fotografica del gas compreso nel piccolo intervallo di velocità tra più 16,4 e più 18,4 chilometri al secondo; l'illustrazione in basso è la sintesi fotografica del gas nell'intervallo di velocità tra più 3,7 e più 5,7 chilometri al secondo. Si noti come l'anello ellittico sia più grande nell'illustrazione in basso, che è stata fatta a velocità minore, che non in quella in alto.**

**Un guscio di gas che si espande radialmente nello spazio interstellare produrrebbe nelle sintesi fotografiche nella pagina a fronte proprio una struttura ellittica o circolare, del tipo di quella qui raffigurata. Qui è rappresentato in sezione un guscio sferico di gas che si espande radialmente in tutte le direzioni (*frecce in nero*). Lo spostamento Doppler, per mezzo del quale l'osservatore misura la velocità del guscio, dipende solo dalla componente della velocità del guscio lungo la linea di vista (*frecce in grigio e in colore*). L'osservatore lontano, esaminando un guscio di questo genere soltanto in un piccolo intervallo di velocità, vedrà un anello le cui dimensioni dipendono dalla velocità lungo la linea di vista. La velocità misurata sarà massima nel punto più vicino (*grigio più scuro*) e nel punto più lontano (*colore più intenso*) del guscio, che sono rispettivamente in avvicinamento e in allontanamento quasi diretto, relativamente all'osservatore; in corrispondenza a quella velocità elevata il diametro misurato del guscio sarà più piccolo. La velocità misurata sarà minima per le aree che si trovano al lembo del guscio (*grigio più chiaro e colore meno intenso*), che si muovono quasi perpendicolarmente rispetto alla linea di vista dell'osservatore; in corrispondenza a tale ridotta velocità il diametro misurato del guscio sarà grande.**

### *I filamenti sono gusci*

In realtà il guscio che si osserva nella sintesi fotografica non è una struttura semplice e ideale. A una qualsiasi velocità l'anello non è circolare, ma ellittico. Il guscio non è visibile in tutto l'intervallo di velocità, ma appare meglio a velocità positive, che corrispondono alla velocità con cui il gas si sta allontanando nella parte più lontana del guscio. Inoltre, nella maggior parte degli intervalli di velocità si può vedere solo una parte del guscio. Va ricordato, però, che il mezzo interstellare non è omogeneo e che pertanto non ci si può attendere un accordo perfetto tra un modello ideale e la realtà. Le deviazioni dai modelli ideali sono così frequenti nelle osservazioni del gas interstellare che in effetti ci riteniamo fortunati per aver osservato un esempio tanto efficace con cui illustrare il modello di un guscio in espansione radiale.

Se si potesse determinare la distanza che ci separa dal guscio ellittico, si potrebbero ricavare molte informazioni sulle proprietà fisiche del guscio stesso. Sappiamo poco sulla distanza di questa particolare struttura, ma i dati disponibili indicano che una distanza di 150 parsec è probabilmente corretta entro un fattore due. A tale distanza le dimensioni angolari del guscio corrispondono a un diametro di 100 parsec e indicano che nel corso della sua espansione il guscio ha spostato una massa di materia interstellare pari a 73 000 masse solari.

Si conoscono due meccanismi che possono dare luogo a gusci di gas interstellare in espansione di così grandi dimensioni. Un meccanismo, proposto recentemente da John I. Castor, Richard Mc Cray e Robert Weaver dell'Università del Colorado, si basa sul trasferimento di energia da un vento stellare (rapido flusso di materia emesso da una stella) altamente energetico. I venti stellari sono stati osservati effettivamente in certi tipi di stelle. Tali stelle non sono molto comuni e non ci sono stelle all'interno del guscio ellittico di gas osservato per cui sia nota la presenza di un vento energetico. È possibile che il guscio sia stato generato da una stella di quel genere che si sia poi evoluta estinguendosi, ma l'energia del guscio è probabilmente maggiore di quella che potrebbe essere fornita da una o più stelle di questo tipo.

L'altro meccanismo capace di produrre un guscio di gas in espansione è una esplosione di supernova; la quasi-distruzione catastrofica di una stella massiccia verso la fine del suo ciclo evolutivo. L'Anello del Cigno, ovvero la Nebulosa del Velo nella costellazione del Cigno, è un involucro in espansione prodotto da una supernova che esplose circa 45 000 anni fa. Si sta espandendo con una velocità di circa 100 chilometri al secondo e ha un raggio di 20 parsec circa. Il nostro guscio ellittico recentemente scoperto è più vecchio di quell'oggetto, ha spazzato via più materia interstellare e si muove più lentamente. Sulla base dei modelli teorici di vecchi gusci di supernova calcolati da Roger A. Chevalier del Kitt Peak National Observatory e William C.

Straka dell'Università di Boston sembra evidente che il nostro guscio ellittico sia stato espulso da una supernova che esplose circa 650 000 anni fa liberando un'energia totale di $4 \times 10^{51}$ erg. Questa quantità di energia è prossima al limite superiore che si accetta generalmente per l'emissione di energia da una supernova e supera probabilmente di molte volte la quantità di energia liberatasi nella formazione dell'Anello del Cigno.

Un guscio di gas interstellare in espansione dovrebbe avere una struttura caratteristica se è stato generato da un vento stellare o da una supernova. Subito dietro il bordo anteriore del guscio il gas è compresso e riscaldato. Via via che questo gas riscaldato è sospinto dall'espansione, si ritira lentamente lasciando il fronte anteriore del guscio e muovendosi verso l'interno. Così facendo si raffredda, ma viene ulteriormente compresso. Se il guscio è il risultato di un vento stellare energetico, l'emissione ultravioletta dalla stella ionizzerà la regione densa dietro al bordo anteriore.

Il guscio è composto da quasi tutta la materia che si trovava originariamente nel volume che è ora occupato dal guscio stesso. L'interno del guscio non è vuoto: contiene un piccolo quantitativo del gas interstellare iniziale, che si trova ora a temperatura molto elevata, bassa densità e alta pressione. Questo gas rarefatto è la sorgente dell'emissione X che si osserva generalmente dall'interno dei resti di supernova (si veda l'articolo *Raggi X da resti di supernova* di Philip A. Charles e J. Leonard Culhane in «Le Scienze», n. 92, aprile 1976). Il gas rarefatto è pure sorgente probabile di ossigeno così altamente ionizzato da essere privato di cinque elettroni. Questa forma altamente ionizzata di ossigeno, indicata con O VI, compare spesso negli spettri ultravioletti delle stelle a elevata temperatura osservate con il telescopio posto a bordo del satellite *Copernicus*. Sia l'emissione ultravioletta da O VI sia l'emissione X sono state osservate nell'interno del nostro guscio ellittico.

L'Anello del Cigno, fotografato nel visibile con il telescopio Schmidt da 122 centimetri di Monte Palomar, è un guscio simile al guscio ellittico che compare nelle sintesi fotografiche di pagina 76. Come è possibile vedere in questa stampa in negativo, l'Anello del Cigno non possiede una forma perfettamente circolare a causa della disomogeneità e della irregolarità della distribuzione del mezzo interstellare circostante. Il guscio si va espandendo con una velocità di circa 100 chilometri al secondo e presenta un raggio di circa 20 parsec. Probabilmente fu generato circa 45 000 anni fa da una supernova, cioè da una stella che esplose catastroficamente.

### *La struttura dei gusci*

Il modo con cui la materia è concentrata in un guscio di tal genere suggerisce la ragione dell'allineamento dei filamenti del gas interstellare con il campo magnetico galattico. Anche il gas più freddo nello spazio interstellare è leggermente ionizzato dalla luce stellare. Dato che gli ioni sono particelle elettricamente cariche, essi fluiranno come una corrente elettrica in presenza di un campo magnetico. Pertanto il gas e il campo magnetico sono legati insieme indissolubilmente dalla debole conduttività elettrica del gas. Man mano che il fronte d'urto del guscio si espande, esso spazza via anche il gas interstellare e anche il campo magnetico. Il campo magnetico dovrebbe essere più intenso nel guscio che nel mezzo interstellare circostante con la stessa proporzionalità con cui la densità del gas nel guscio è più elevata che nel mezzo circostante. Inoltre, la direzione del campo magnetico tenderà a essere parallela alla superficie del guscio.

Non è ben chiaro come si formino i filamenti dentro a un guscio. È possibile che si generino con lo stesso processo che crea i filamenti nella corona solare. Una gran quantità di lunghi filamenti è visibile in tutte le sintesi fotografiche del mezzo interstellare a piccoli intervalli di velocità; ce ne sono molti di più di quelli distinguibili sulle fotografie che ho presentato in questo articolo. La maggior parte dei filamenti non mostra cambiamenti di struttura al variare della velocità. Ciononostante i miei colleghi e io riteniamo che anche i filamenti siano parti del guscio in espansione, parti rallentate da collisioni con il gas interstellare circostante o con altri gusci.

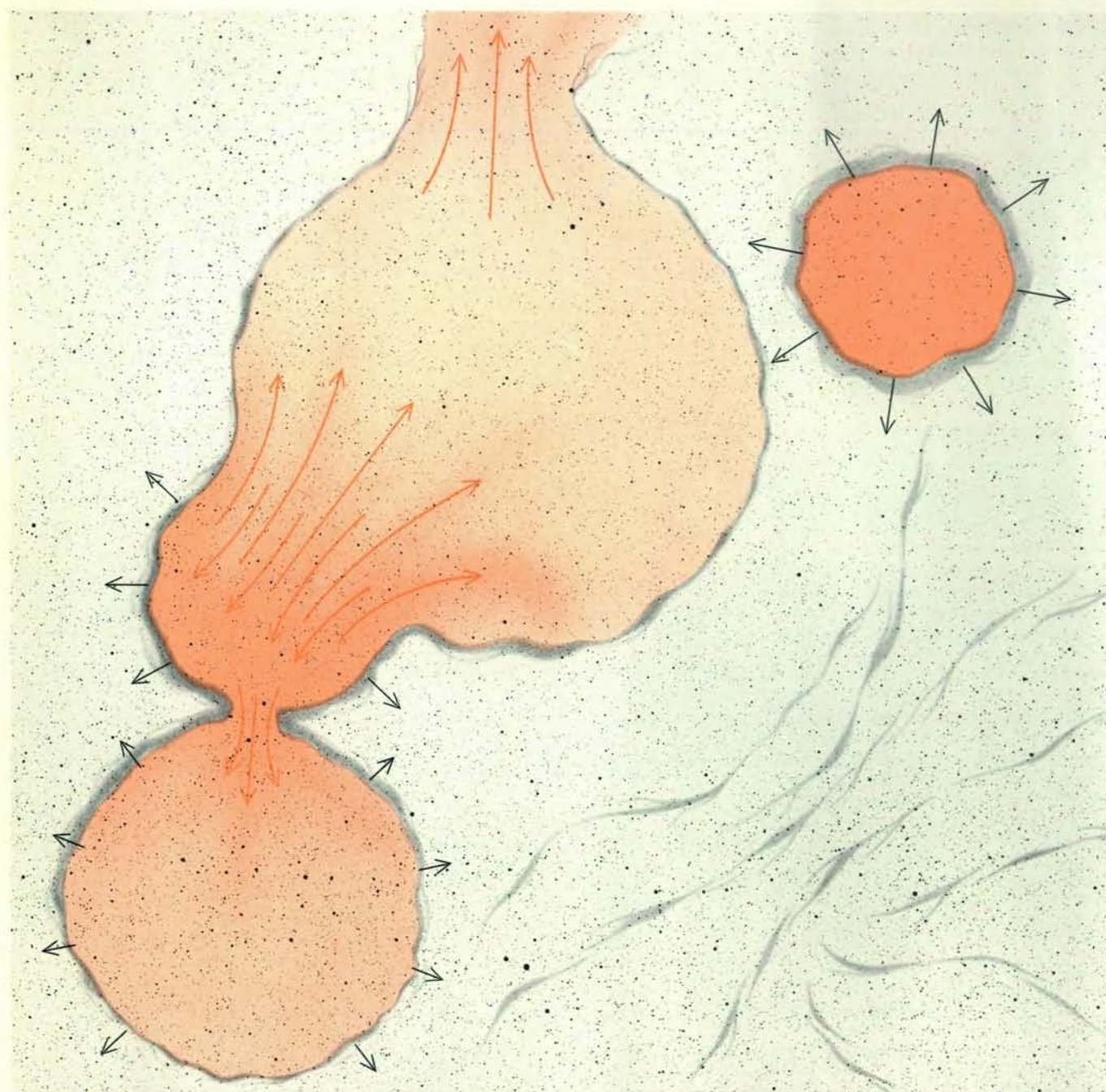
La nostra interpretazione è un'estrapolazione basata sulle osservazioni relativamente affidabili dei pochi gusci che presentano dimensioni diverse in intervalli di velocità diversi, ma pensiamo che sia giustificata. Una ragione è connessa al fatto che nel caso di un guscio che presenta una pronunciata variazione di diametro con la velocità e nel caso di due altri gusci che mostrano variazioni di diametro scarse o nulle con la velocità si hanno prove indipendenti, sotto forma di emissione di sincrotrone nel radio, che le strutture sono conseguenza di un'esplosione di supernova. L'emissione di sincrotrone è prodotta da elettroni di alta energia che si muovono a spirale con velocità prossime a quella della luce in un campo magnetico interstellare. Questo tipo di emissione è spesso associato a gusci di supernova; si ritiene che gli elettroni siano stati accelerati a energie elevate nella esplosione della supernova.

I numerosi filamenti visibili nelle sintesi fotografiche implicano la presenza di molti gusci nello spazio interstellare. La esistenza di un numero di gusci così elevato ha conseguenze interessanti che sono state studiate teoricamente da Donald P. Cox e Barham W. Smith dell'Università del Wisconsin. Se vi sono molti gusci nello spazio interstellare, è probabile che entrino in collisione espandendosi. Dopo la collisione le regioni interne dei gusci risultano connesse tra loro e il gas a elevata temperatura che si trova dentro ai gusci è libero di scorrere rapidamente da un guscio all'altro. È probabile che in una situazione stazionaria vari gusci si uniscano formando tunnel di gas caldo a bassa densità, circondati da involucri sottili di gas freddo più denso e da campi magnetici. Talvolta un guscio potrebbe scoppiare come una bolla nel caso in cui la materia più fredda non fosse sufficiente per contenere la pressione del gas a elevata temperatura. Questo fornirebbe una via di fuga per il gas caldo confinato, che potrebbe così uscire fuori dallo spazio limitato dai gusci.

### *La descrizione a gusci*

Il rilevamento da noi effettuato nel radio sembra coerente con questa descrizione della struttura del mezzo interstellare; tale descrizione sembra confermata anche dalle osservazioni di raggi X e dell'emissione ultravioletta dell'O VI nello spazio interstellare. La descrizione a gusci del mezzo interstellare differisce radicalmente dal modello a nubi e non è ancora generalmente accettata. Rimangono problemi importanti a cui si dovrà dare una risposta prima che questa descrizione possa essere considerata del tutto corretta. Le osservazioni nell'ottico sono in accordo con questa descrizione oppure forniscono dati su componenti ancora diverse del gas interstellare? La temperatura e la distribuzione di densità del gas interstellare osservate sono in accordo con ciò che ci si aspetterebbe di trovare dietro al bordo anteriore del guscio? La topologia e l'intensità dei campi magnetici nello spazio interstellare sono coerenti con questa interpretazione?

Nuovi strumenti appena completati o in costruzione avranno una funzione importante nello sforzo volto a risolvere questi problemi. In particolare il radiotelescopio da 300 metri di Arecibo a Portorico di cui è stata rifatta recentemente la superficie e il grande insieme di radiotelescopi nel New Mexico dovrebbero fornire molti dei dati radio necessari. Occorrono dati sia nella banda X sia nell'ultravioletto per determinare se il gas ad alta temperatura è associato preferenzialmente o no all'interno dei gusci. Questi dati si possono ottenere solo con strumenti portati nello spazio. Intanto gli astronomi stanno osservando da terra il mezzo interstellare nel visibile con nuovi strumenti capaci di fornire dati di qualità molto più elevata, meglio confrontabili con i dati ottenuti nel radio.

Vi sono molti problemi sull'origine dei gusci. L'energia necessaria per creare tali gusci è grande, forse più grande dell'energia che può essere fornita dai venti stellari o dalle esplosioni di supernova. Il numero di gusci sembra grande. Anche se nessuno finora ha analizzato in dettaglio questa molteplicità, è possibile che ci siano troppi gusci rispetto alla frequenza generalmente ammessa con cui si formano stelle dotate di venti stellari energetici o con cui esplodono supernove. Se è così, forse le stelle con venti stellari energetici o le supernove sono più frequenti di quanto crediamo ora. O forse c'è un altro tipo di oggetto, finora sconosciuto, che immette enormi quantità di energia nel mezzo interstellare.

**La descrizione a gusci del mezzo interstellare è proposta dall'autore per spiegare la struttura filamentare del mezzo interstellare. Un guscio di gas è prodotto da un vento stellare energetico emesso da una stella calda o a seguito dell'esplosione di una supernova. Un guscio di recente formazione si espande rapidamente (*in alto a destra*), accumulando il gas che si trova nel mezzo interstellare freddo circostante (*in grigio chiaro*) e comprimendolo e riscaldandolo così da formare un denso bordo esterno di gas ionizzato analogo a quello visibile nell'Anello del Cigno (*in grigio scuro*). Man mano che il guscio si espande lascia dietro nel suo interno gas ad alta temperatura e a bassa densità (*in colore*) che emette raggi X e produce inoltre negli spettri ultravioletti delle stelle calde le righe di assorbimento proprie dell'ossigeno altamente ionizzato. Mentre il guscio diventa più grande e più vecchio, il gas viene ulteriormente compresso e raffreddato. Gli elettroni si ricombinano con gli ioni formando idrogeno atomico, che emette nel radio a 21 centimetri. Nel frattempo l'espansione del guscio rallenta (*in basso a sinistra*). Secondo il lavoro teorico di Donald P. Cox e Barham W. Smith dell'Università del Wisconsin, la coesistenza di un numero notevole di grandi gusci in espansione comporterà collisioni tra loro. Quando ciò accade i loro interni risultano connessi, così che il gas caldo può passare da un guscio a un altro. Se i gusci in connessione saranno in numero sufficiente, il gas caldo potrà scorrere attraverso tunnel nel mezzo interstellare formato da gusci freddi e densi e da campi magnetici. Gusci molto vecchi potranno infine diventare così sottili che la loro parte più debole potrà scoppiare come una bolla (*in alto al centro*), permettendo in tal modo al gas caldo di sfuggire nello spazio circostante. Resti di gusci antichi e ormai informi danno luogo a strutture filamentari minori (*in basso a destra*).**

# Le formiche tessitrici

*Questi insetti, il cui comportamento è coordinato da stimoli chimici complessi, usano le larve come spole per cucire le foglie con cui costruiscono i nidi nelle foreste pluviali dell'Africa e dell'Asia*

di Berthold K. Hölldobler ed Edward O. Wilson

Alcune specie d'insetti vivono in un ordine sociale progredito, caratterizzato dalla cooperazione, dalla specializzazione delle caste e dall'altruismo individuale. Tra le migliaia di specie di insetti sociali alcune meritano d'essere chiamate classiche, poiché certe caratteristiche notevoli nel loro comportamento hanno stimolato studi straordinariamente accurati e completi. Le api, i bombi, le formiche scacciatrici, le formiche legionarie, le formiche taglia-foglie, quelle schiaviste e le termiti coltivatrici di funghi, si possono considerare tutti esempi classici di insetti sociali. Gli ultimi animali di questo tipo a essere esaminati sono state le formiche tessitrici del genere *Oecophylla*, che vivono in Africa e nell'Asia tropicale. Queste formiche dedicano gran parte del loro repertorio di comportamento alla comunicazione, che viene ulteriormente arricchita da variazioni basate sul sistema di casta: una colonia di formiche tessitrici riesce di conseguenza a compiere azioni straordinarie, che vanno molto al di là della capacità delle singole formiche.

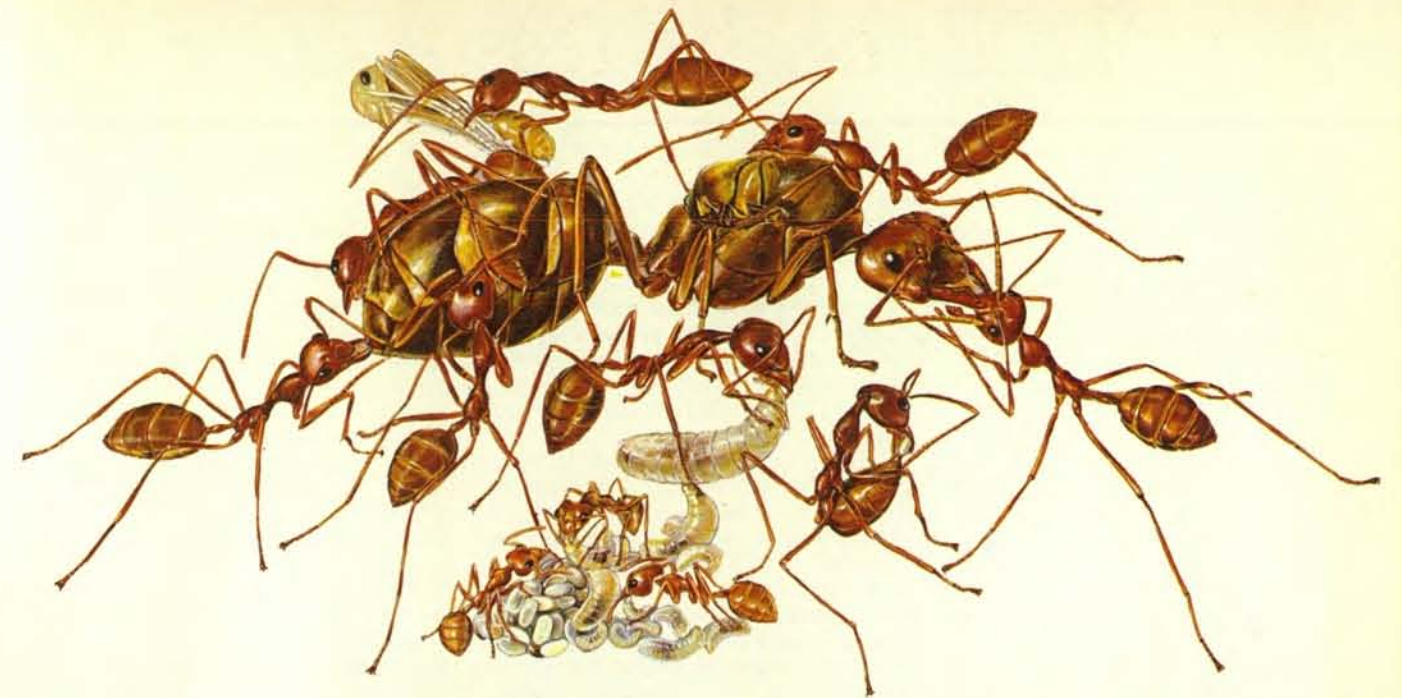
Le formiche tessitrici, insetti territoriali e aggressivi, sono assai numerose e hanno acquisito una posizione d'importanza ecologica eccezionale nelle foreste pluviali, nelle piantagioni di cacao e negli ambienti boschivi del medesimo tipo, loro habitat normale; perciò sono state compiute osservazioni sempre più numerose e approfondite su questi insetti. Negli ultimi due anni siamo riusciti anche ad allevare in laboratorio colonie della specie africana *Oecophylla longinoda*, inducendo le formiche a vivere in alberi piantati in vaso e in tubi di vetro: in queste condizioni è stato possibile per la prima volta studiare particolareggiatamente l'intera gamma di comportamenti che caratterizzano la vita sociale delle formiche tessitrici.

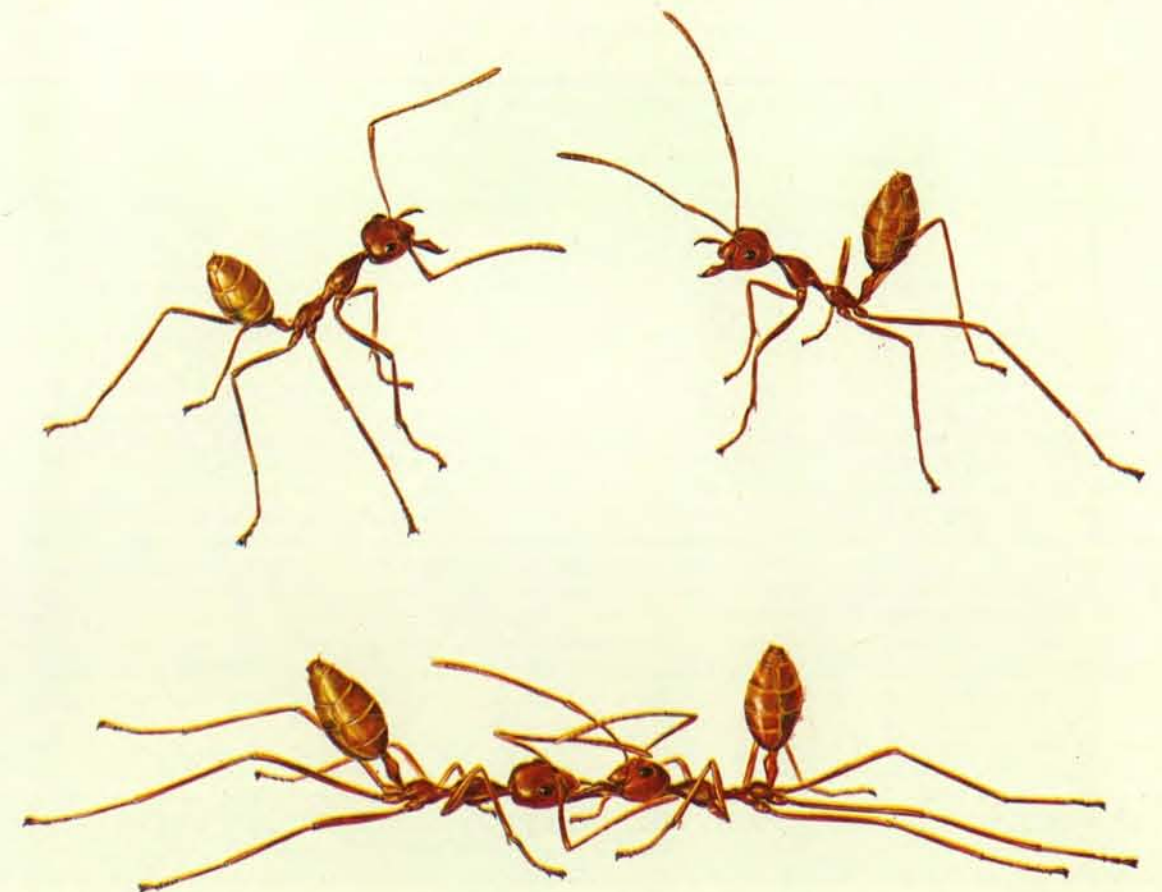
Questi snelli insetti di color giallo sono perfettamente adattati alla vita delle foreste tropicali. La loro principale unità sociale è la colonia, che è costituita da circa 500 000 femmine operaie, progenie di un'unica enorme regina.

Il sistema di caste all'interno di ciascuna colonia è costituito da tre forme di femmine adulte: la regina, dal corpo grande e pesante, una numerosa popolazione di operaie «maggiori» e una popolazione meno numerosa di operaie «minori». I maschi delle formiche tessitrici, come quelli di altre specie di formiche, partecipano relativamente poco alla vita sociale della colonia: subito dopo aver raggiunto la maturità sessuale lasciano il nido per partecipare ai voli nuziali con le regine vergini, dopo di che, oramai esausti, muoiono, senza aver fatto ritorno al nido.

Le operaie maggiori sono di dimensioni piuttosto grandi: in media hanno una lunghezza di 6 millimetri. Adempiono le funzioni di operaie generiche, poiché sono responsabili della maggior mole di lavoro nella ricerca del cibo e nella costruzione del nido; più aggressive che non le operaie minori, si precipitano fuori dal nido al minimo disturbo, per mordere gli intrusi e secernere acido formico con le loro ghiandole velenifere. Inoltre esse costituiscono una scorta fittamente disposta attorno alla regina. La afferrano strettamente con le zampe potenti, tanto, a volte, da tenerla sollevata a mezz'aria nel centro della cavità del nido. Ogni minuto circa una delle operaie maggiori rigurgita un cibo liquido nella bocca della regina. A intervalli un po' meno frequenti un membro della scorta depone uno speciale uovo trofico, un oggetto flaccido senza alcuna possibilità di sopravvivere, che viene immediatamente fornito come cibo alla regina. Grazie a questo tipo di alimentazione ininterrotta la regina riesce a sviluppare centinaia di uova al giorno. Non appena le uova escono dall'ovidotto, le operaie maggiori le sistemano in speciali mucchi d'incubazione. Qui le operaie minori, di dimensioni più piccole, prima si prendono cura delle uova, poi nutrono e lavano le minuscole larve che ne sgusciano. Quando le larve si approssimano alle dimensioni massime, le operaie maggiori e le operaie minori si dividono praticamente a metà il compito di curarle.

Le formiche tessitrici devono il loro nome al metodo seguito nella costruzione del nido. I nidi vengono fabbricati con foglie piegate e legate assieme in modo da formare compartimenti ben chiusi, simili a tende. Le foglie sono tenute a posto da cuciture di seta filata dalle larve, che le operaie maggiori utilizzano come spole per tessere i nidi. Questo modo di costruzione del nido è uno degli esempi più notevoli che si conoscano di cooperazione sociale esistente tra gli animali inferiori.

Non appena le formiche tessitrici hanno scelto un ramo d'albero adatto per nidificare, sciamano sulle foglie di quel ramo e iniziano a tirarne le punte e i bordi. Quando una formica riesce a voltare verso l'alto un pezzetto di foglia, le operaie vicine vengono attirate verso quella parte e presto si forma un gruppetto di formiche che tirano all'unisono. Quando una foglia è più larga rispetto alla lunghezza del corpo d'una formica, o quando due foglie devono essere avvicinate attraverso uno spazio vuoto, le operaie formano ponti viventi tra i due punti che debbono essere congiunti. Poi alcune formiche della catena si arrampicano sul dorso delle vicine e tirano all'indietro: in tal modo accorciano la lunghezza della catena e avvicinano i bordi delle foglie. Quando le foglie sono state manipolate in modo da assumere la forma voluta, alcune formiche rimangono sopra di esse e usano le zampe e le mandibole per tenerle a posto. Altre formiche si recano nei nidi già costruiti e ritornano poi verso il sito di nuova costruzione portando con sè alcune larve già parzialmente cresciute. Le operaie fanno muovere le larve avanti e indietro, trasversalmente alle linee di giunzione delle foglie: questo movimento induce le larve a secernere filamenti di seta dalle aperture ghiandolari situate appena sotto la bocca. Migliaia di questi fili intrecciati in modo da formare un tessuto risultano abbastanza tenaci da fissare nella posizione voluta le foglie. Allo stesso modo vengono poi tessuti fogli serici, in modo da formare entrate circolari e gallerie



Il sistema di caste della formica tessitrice africana è costituito da tre forme di femmine adulte: una sola regina, una popolazione di operaie «maggiori» (di grandi dimensioni) che vanno in cerca di cibo ed eseguono molti altri compiti, e un numero inferiore di operaie «minori» (di piccole dimensioni) che si prendono cura delle uova e delle larve più giovani. La regina (*al centro*) trascorre il suo tempo solamente nel mangiare cibo, di solito rigurgitato, fornitole dalle operaie maggiori e nel deporre uova. In primo piano vediamo un'operaia maggiore mentre fornisce cibo rigurgitato a una larva; una seconda formica operaia depone un uovo «trofico», privo, cioè, di capacità di sopravvivenza, che verrà poi offerto alla regina come cibo speciale. Le operaie minori sono raffigurate, in basso al centro, raggruppate attorno a un mucchio di uova in incubazione. La scena è stata disegnata da Turid Hölldobler: le formiche sono ingrandite di circa 6 volte.

Il combattimento territoriale di due formiche tessitrici inizia con una specie di danza, in cui le contendenti si rizzano estendendo le zampe e girano l'una attorno all'altra con movimenti rigidi, a scatti (*in alto*). Poi una tenta di afferrare l'altra con le mandibole (*in basso*).

esterne che conducono all'interno del nuovo nido.

Una singola colonia di formiche tessitrici può occupare un volume notevole nella volta d'una foresta. La colonia può occupare tutto un albero o addirittura più alberi adiacenti, senza che vengano a interrompersi le linee di comunicazione d'importanza vitale per gli insetti sociali. Dalle foglie le formiche tessitrici ricavano centinaia di nidi che funzionano da ricoveri, da nidi d'infanzia, da avamposti. Durante il giorno le operaie foraggiatrici, deputate alla ricerca del cibo, setacciano ogni centimetro quadrato di foglie e di corteccia all'interno del loro territorio. Sbaragliano i nemici, catturano insetti e raccolgono la melata, l'escremento dolce di sciami di afidi e altri omotteri che si nutrono di linfa e a cui le formiche accudiscono come se fossero vacche da latte.

Le formiche del genere *Oecophylla* non sono i soli insetti che tessono. Alcune altre specie tropicali utilizzano la seta larvale per costruire nidi sugli alberi e fra i cespugli. Tuttavia le formiche tessitrici si distinguono per il loro stretto controllo sopra l'ambiente circostante, al punto che l'uomo le ha addirittura utilizzate per regolare l'ambiente arboreo a proprio vantaggio.

Secondo documenti storici, i nidi delle formiche tessitrici nella zona attorno a Canton, in Cina, furono raccolti, venduti e posti su alberi scelti di agrumi attorno all'anno 300 d.C. per combattere gli insetti parassiti. La medesima tecnica fu applicata nel dodicesimo secolo ed era ancora praticata nella Cina meridionale in pieno ventesimo secolo. La specie utilizzata a tale scopo è stata la *Oecophylla smaragdina*, una specie asiatica. Questa utilizzazione di formiche tessitrici costituisce l'esempio più antico, fra quelli noti, di controllo biologico degli insetti nella storia dell'agricoltura.

Recentemente Dennis Leston, già dell'Università del Ghana, e altri entomologi hanno consigliato l'impiego della specie africana di formiche tessitrici per combattere gli insetti nocivi di importanti piante arboree coltivate come il cacao. Gli studi compiuti in Ghana hanno messo in rilievo che la presenza di formiche tessitrici riduce l'incidenza di due delle più gravi malattie del cacao, provocate rispettivamente da un virus e da un fungo. In entrambi i casi il germe patogeno è trasmesso da alcune cimici delle foglie: evidentemente le formiche tessitrici, distruggendo le cimici, combattono le malattie delle piante di cacao. Le operaie del genere *Oecophylla* sono anche particolarmente abili nell'attaccare gli insetti che sono soliti cibarsi dei tessuti e della linfa degli alberi.

L'eccezionale capacità di controllo ambientale delle formiche tessitrici è stata ottenuta grazie all'evoluzione di forme avanzate di comportamento sociale. Il sistema di comunicazione che abbiamo osservato nei nostri studi sulla specie africana di formiche tessitrici è fra i più complessi e progrediti di cui si abbia notizia tra gli insetti sociali. La grande peculiarità delle formiche tessitrici consiste nella loro capacità, dimostrata nella costruzione del nido, di cooperare in attività di gruppo.

Questi insetti impiegano cinque differenti sistemi di reclutamento, che consistono in combinazioni distinte di segnali chimici e tattili, per compiere lavori in cui è necessaria una collaborazione di gruppo. Questi sistemi di reclutamento sono impiegati nelle principali occupazioni delle formiche tessitrici fuori del nido: la penetrazione di un nuovo territorio, la difesa di tale territorio e la ricerca e lo sfruttamento delle fonti di cibo nella nuova area.

Le formiche tessitrici possiedono un impressionante senso dell'orientamento, che serve loro per assicurarsi nuovo territorio. Grazie ai grandi occhi, sono fornite di una vista insolitamente acuta tra le formiche. Inoltre sono capaci di ricordare l'aspetto di molti particolari dell'area del nido. Se un oggetto viene semplicemente spostato da un lato dell'albero che costituisce il nido a un altro lato del medesimo albero, le operaie escono dal nido per esplorare l'oggetto, come se fosse un territorio nuovo, da colonizzare. In realtà, quando un oggetto di cospicue dimensioni, come una pianta in vaso o una scatola, viene spostato vicino a un nido di formiche tessitrici, le attente operaie escono e strisciano lungo

**Una catena vivente di formiche tessitrici sta piegando verso l'alto la punta d'una foglia durante la costruzione d'un nido arboreo. Le operaie accorciano le catene arrampicandosi sul dorso delle compagne e tirando contemporaneamente il bordo della foglia dietro di sè.**

i rami e le foglie del loro albero, nel tentativo di raggiungerlo. Se non riescono a raggiungere la nuova superficie con le proprie zampe, cominciano ad arrampicarsi una sopra l'altra, costruendo piramidi o catene coi loro corpi finché sull'abisso viene gettato un ponte vivente. A quel punto le operaie si precipitano sopra il nuovo territorio e cominciano a esplorarlo.

Le prime esploratrici ritornano al nido per reclutare altre operaie che le aiutino a rendere sicuro il nuovo territorio. Contrassegnano la via dal territorio al nido con una pista odorosa, ossia con tracce d'una sostanza chimica, un ferormone, portatrice d'un messaggio che guida le compagne di nido verso la nuova area. Le tracce odorose vengono create in un modo insolito. Le formiche fanno protendere verso l'esterno un segmento ghiandolare dell'intestino posteriore attraverso l'ano: questo organo, che chiamiamo ghiandola rettale, e la cui esistenza è nota solo nelle formiche tessitrici, viene usato in quattro dei sistemi di reclutamento. Quando la ghiandola è in posizione estroflessa rimane appoggiata sopra una minuscola slitta costituita da due setole che sporgono dalla punta dell'addome. Quando la formica ritorna correndo verso il nido, sparge il secreto della ghiandola rettale sul terreno, creando così la pista odorosa. Quando questa formica incontra le compagne di nido, incurva il corpo nella loro direzione, toccandole sulla testa con le antenne. Il «saluto» fornisce uno stimolo alle compagne, tanto che queste sono indotte a seguire la traccia verso il nuovo territorio e cominciano a esplorarlo. Possiamo chiamare questo tipo di processo «sistema di reclutamento verso un nuovo territorio».

Dopo che è stata assicurata questa nuova estensione del territorio della colonia, le operaie la setacciano in lungo e in largo in cerca di cibo. Quando una operaia s'imbatte in una secrezione zuccherina (di solito emessa da afidi), ritorna al nido per reclutare compagne. Ancora una volta utilizza la ghiandola rettale per lasciare una traccia odorosa. In questo sistema di reclutamento l'operaia stimola le sue compagne accarezzandole con le antenne e offrendo loro il cibo zuccherino trovato, per mezzo del rigurgito. L'emigrazione verso i nidi costruiti nel nuovo territorio viene effettuata impiegando un sistema di reclutamento ancora diverso, che comprende la creazione d'una traccia odorosa mediante la secrezione della ghiandola rettale, segnali di natura tattile (le carezze con le antenne) e addirittura il trasporto fisico delle compagne di nido.

**In questa sequenza di disegni è raffigurata la costruzione del nido delle formiche tessitrici che lavorano in cooperazione. Le operaie iniziano il lavoro indipendentemente, nel tentativo di tirare verso il basso o arrotolare verso l'alto le foglie. Quando gli sforzi di una o più formiche ottengono qualche risultato, in una parte qualsiasi della foglia, altre operaie nelle vicinanze abbandonano i propri tentativi e partecipano allo sforzo comune (*a*). Quando alla fine le foglie hanno assunto la forma di tende da campo, alcune formiche continuano a tenere ferma la struttura con le zampe e le mandibole, mentre altre portano larve parzialmente cresciute (*in colore*) dai nidi preesistenti e legano insieme le foglie con seta larvale, appiccicosa (*b*). In seguito vengono aggiunti fogli di seta con i quali vengono formate entrate circolari e gallerie (*c*).**

Due altri sistemi di reclutamento vengono utilizzati per difendere il territorio della colonia. Le formiche tessitrici sono particolarmente aggressive contro i membri di altre colonie di formiche tessitrici. In effetti, Leston ha trovato che i territori di colonie differenti sono separati da «terre di nessuno», fasce ri-

La seta per legare i nidi è filata da larve, che le operaie tengono nelle mandibole e muovono avanti e indietro sopra la giuntura delle foglie, come se fossero spolette. I filamenti, secreti da aperture ghiandolari situate appena sotto la bocca della larva, vengono tessuti in teli.

Nido costruito da una colonia di formiche tessitrici africane in un albero di pompelmo piantato in vaso nel laboratorio degli autori, all'Università di Harvard. Allo stato selvatico una colonia di circa mezzo milione di operaie (progenie d'un unica regina) può costruire centinaia di nidi con le foglie di uno o più alberi. Le formiche perlustrano questo dominio di giorno e si ritirano di notte all'interno dei nidi.



strette in cui poche formiche si avventurano. Il contatto tra colonie diverse di formiche tessitrici di solito si risolve in una guerra immediata e spettacolare, che provoca vittime numerose e si conclude con la ritirata di una delle colonie da tutto il proprio territorio o almeno da una parte di esso. Nell'habitat naturale della specie africana queste battaglie possono durare anche per diversi giorni, e le contendenti, ammassate, continuano a lottare, spostando lentamente e progressivamente le linee difensive.

Le operaie foraggiatrici che incontrano formiche tessitrici nemiche reagiscono con una serie di movimenti rapidi e precisi. Il combattimento individuale inizia con una manovra simile a una danza in cui le contendenti si rizzano sulle zampe e girano con movimenti rigidi, a scatti. Poi si lanciano l'una contro l'altra e cercano di mordersi con le mandibole. La formica sconfitta viene inchiodata a terra con le zampe divaricate; poi le vengono amputate le zampe e le antenne, e spesso le viene squarciato l'addome. Nella confusione della mischia le formiche uccise o agonizzanti si contorcono. Alcune operaie ritornano precipitosamente verso il nido, lasciando tracce odorose con la ghiandola rettale: quando incontrano le compagne di nido, scattano col corpo in una posizione che sembra una versione ritualizzata della danza precedente il combattimento. Le compagne rispondono, non con la lotta, ma correndo lungo la pista odorosa fino al luogo della battaglia.

Nel medesimo tempo le operaie impegnate nel combattimento usano un sistema di reclutamento a breve raggio per organizzare attacchi in gruppo. Quando una formica foraggiatrice incontra una nemica ma non riesce a impegnarla in un combattimento, spesso percorre velocemente brevi giri, trascinando l'addome sul terreno. In questo sistema di reclutamento la formica ruota il segmento terminale dell'addome per mettere allo scoperto la ghiandola sternale. L'esistenza di quest'organo è stata scoperta recentemente ed è nota solamente nelle formiche tessitrici. Le formiche sono attirate dal profumo del secreto della ghiandola sternale a una distanza massima di 10 centimetri: perciò nel punto in cui l'operaia si è imbattuta con la formica rivale si vengono a formare piccoli grappoli di formiche.

J.W.S. Bradshaw e i suoi colleghi, all'Università di Southampton, hanno dimostrato che questo effetto viene accresciuto dall'emissione di diverse sostanze d'allarme da parte di ghiandole localizzate alla base delle mandibole della formica tessitrice. Alcune di queste sostanze attirano compagne di nido sulla scena, altre aumentano il livello di eccitamento e di aggressività nelle formiche. Abbiamo osservato ripetutamente che questi gruppi di formiche reclutate sono estremamente più efficienti, nel combattimento, rispetto alle formiche individuali. Per esempio, le operaie della grande formica nera degli alberi, *Polyrhachis militaris*,

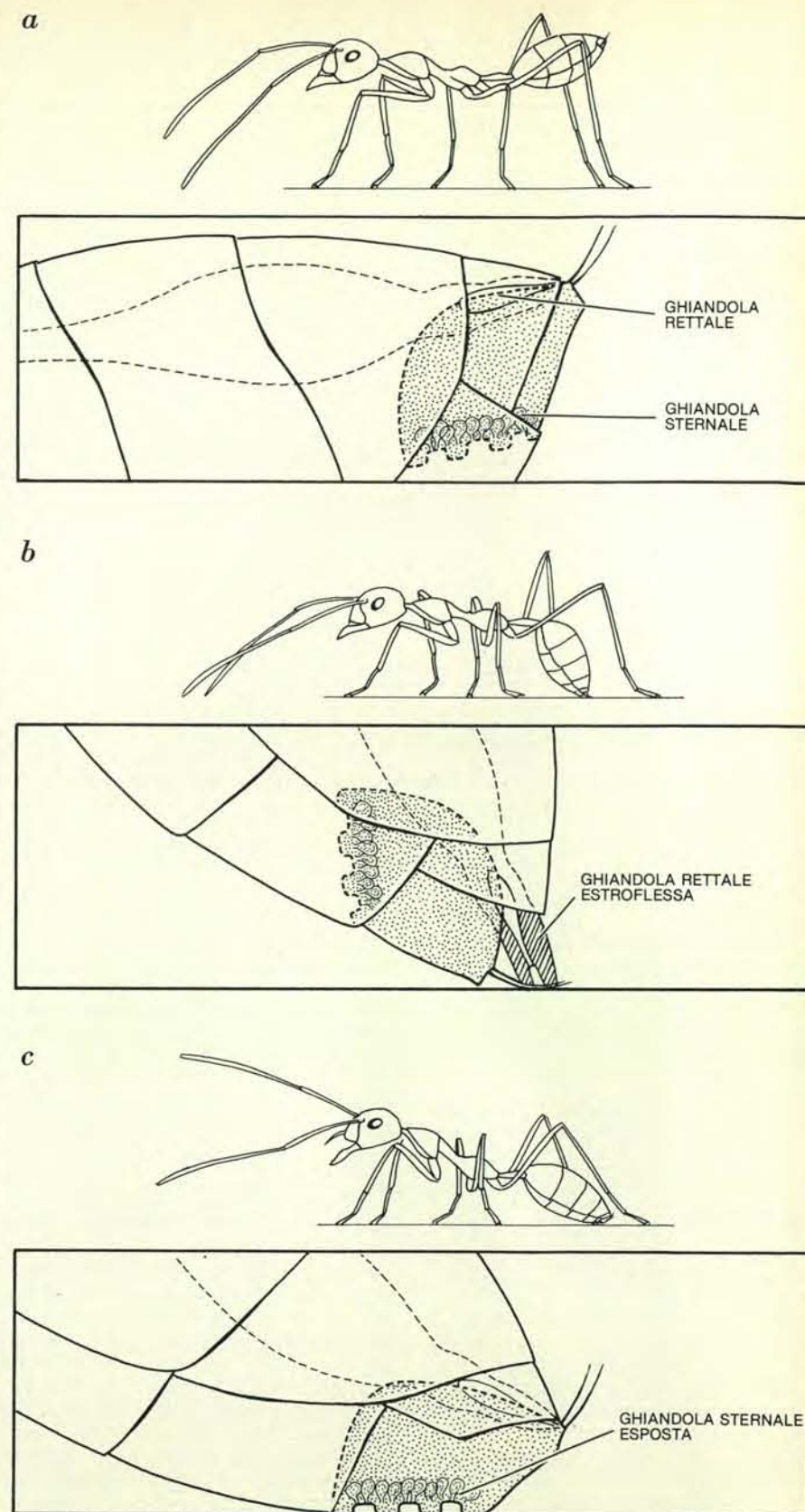
a

GHIANDOLA RETTALE
GHIANDOLA STERNALE

b

GHIANDOLA RETTALE ESTROFLESSA

c

GHIANDOLA STERNALE ESPOSTA

Per coordinare l'attività sociale le formiche tessitrici depositano tracce odorose che formano una pista. Normalmente l'operaia foraggiatrice cammina con l'addome sollevato (*a*). Quando incontra un nuovo terreno o una nuova fonte di cibo, abbassa l'addome, estroflette la ghiandola rettale attraverso l'ano e deposita un feromone, ossia una sostanza portatrice di messaggio, in una linea lungo il terreno (*b*). Quando la ghiandola rettale è estroflessa, rimane appoggiata su una minuscola slitta costituita da due setole. Le compagne di nido che sono state stimolate con un segnale tattile dalla formica che sta marcando la pista ne seguono le tracce fino al nuovo territorio. Il reclutamento a breve raggio eseguito dalle operaie maggiori per combattere gli insetti invasori o altri intrusi viene compiuto mettendo allo scoperto la ghiandola sternale dell'addome e trascinandola sul terreno per depositare una pista odorosa breve, ad anello (*c*).

| SISTEMA | FUNZIONE | SEGNALI CHIMICI | SEGNALI TATTILI | TIPO DI MOVIMENTO |
|---|---|---|---|---|
| Reclutamento per il cibo | Reclutamento delle operaie maggiori verso una fonte di cibo immobile, costituita particolarmente da materiali zuccherini | Traccia odorosa lasciata dalla ghiandola rettale e rigurgito del cibo liquido raccolto | Tocco con le antenne, ondeggiamento della testa e apertura delle mandibole, associati con l'offerta di cibo | Piste sinuose tracciate attorno alla fonte di cibo, con la pista principale diretta verso il nido |
| Reclutamento verso un nuovo territorio | Reclutamento di operaie maggiori per il nuovo territorio | Pista odorosa tracciata dalla ghiandola rettale | Tocco mediante le antenne e di tanto in tanto movimento a scatto del corpo in avanti e indietro | Piste di largo raggio tracciate attorno al nuovo territorio con deposizione di materiale dall'intestino posteriore, contenente un ferormone territoriale; la pista principale porta direttamente verso il nido |
| Reclutamento durante l'emigrazione | Emigrazione di membri della colonia verso una nuova località di nidificazione | Pista odorosa tracciata dalla ghiandola rettale | Toccando con le antenne, una formica indica la propria disponibilità a trasportare un'altra formica nella nuova località di nidificazione | La pista principale conduce direttamente alla località di nidificazione senza aggiungere alcuna traccia secondaria. Le operaie portano dapprima soprattutto larve e pupe, poi altre operaie, verso la località prescelta |
| Reclutamento a breve raggio per combattere i nemici | Reclutamento a breve distanza di compagne di nido per farle raggruppare attorno agli invasori e alle prede, permettendo così una cattura più rapida | Piste odorose sinuose e a breve raggio tracciate mediante la ghiandola sternale; esposizione della superficie ghiandolare quando l'addome viene sollevato in aria | Nessuno | Piste sinuose di breve raggio, tracciate solo in vicinanza o a contatto del nemico |
| Reclutamento a largo raggio per combattere i nemici | Reclutamento di operaie maggiori per combattere gli invasori, particolarmente intenso durante le guerre territoriali contro membri della medesima specie | Pista odorosa lasciata dalla ghiandola rettale | Tocco con le antenne. Durante i periodi di maggiore eccitamento, il corpo si muove a scatti in avanti e indietro | La pista principale conduce direttamente al nido |

**Sono qui riassunti schematicamente i sistemi di reclutamento della formica tessitrice. I due ferormoni secreti dalla ghiandola rettale e da quella sternale, combinati con segnali tattili e con la configurazione spaziale della pista odorosa, possono comunicare cinque diversi messaggi.**

possono mettere facilmente fuori combattimento le formiche tessitrici isolate. Tuttavia, quando tre o più formiche tessitrici formano un gruppo compatto, riescono ad afferrare tutte assieme una formica degli alberi e a inchiodarla al terreno. Altre operaie arrivano rapidamente sul posto e collaborano all'uccisione della nemica.

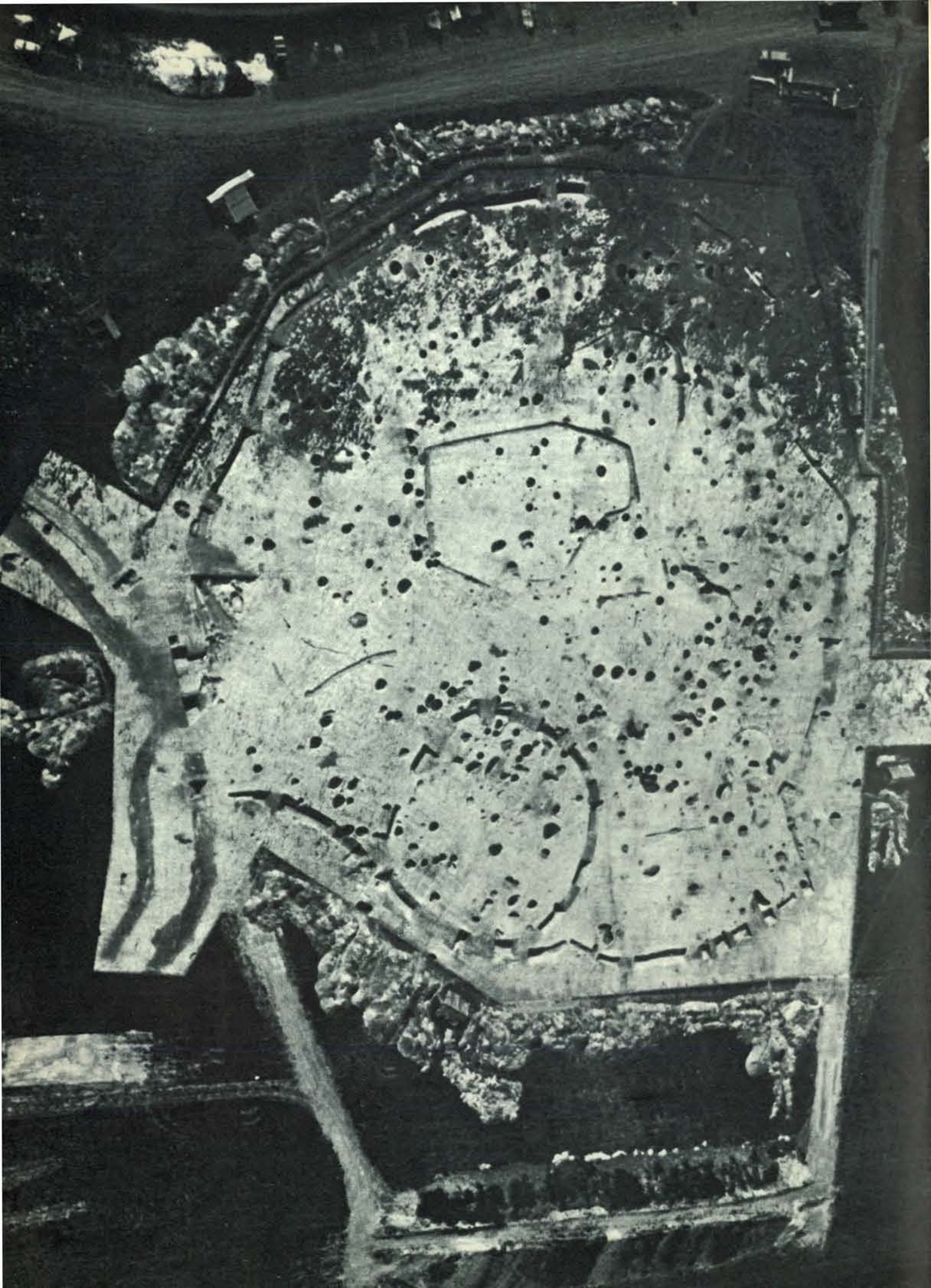
La formica tessitrice è una fine intenditrice di segnali odorosi. Abbiamo notato che in laboratorio le operaie foraggiatrici, dopo aver occupato un nuovo territorio, cominciano a depositare gocciole relativamente grosse di materiale fecale sopra la superficie del territorio. Questo comportamento differisce da quello esibito dalla maggior parte delle altre formiche, che concentrano gli escrementi in località di scarico di rifiuti e in altre zone ristrette. Quando le formiche tessitrici perlustrano il loro territorio, ispezionano le gocciole fecali. Se una formica incontra una gocciola lasciata da un membro di un'altra colonia, reagisce istantaneamente con avversione, assume una posizione ostile e poi esamina più da vicino la gocciola. Siamo stati in grado, nel nostro laboratorio, di indurre la stessa sequenza di reazioni con fluidi raccolti dall'intestino posteriore di formiche estranee.

Le sostanze fecali costituiscono un fattore di vantaggio per le formiche tessitrici, quando si presenta il caso che debbano difendere in combattimento il loro territorio. Abbiamo fatto in modo da provocare una serie di otto «guerre» tra colonie in aree precedentemente contrassegnate con gocciole fecali di formiche dell'una o dell'altra colonia. In ogni caso i membri della colonia che avevano depositato le gocciole erano meno esitanti nella ricerca del cibo sopra il terreno e più rapidi nel reclutare compagne di nido quando incontravano formiche estranee. Di conseguenza si trovavano a poter godere di un vantaggio iniziale e si assicuravano in tal modo un territorio più esteso durante le prime fasi del combattimento.

Le caratteristiche distintive delle formiche tessitrici si sono evolute certamente in tempi assai remoti. Le specie africana e asiatica sono le sopravvissute di uno dei ceppi più tipici e antichi di formiche. Molti insetti fossili degli ultimi 100 milioni d'anni si sono conservati nell'ambra, che è resina fossilizzata. I fossili di due specie estinte di formica tessitrice, *Oecophylla brevinodis* e *Oecophylla brischkei*, si trovano nell'ambra depositata nell'area del Mar Baltico circa 30 milioni d'anni fa, nel periodo oligocenico. Durante tale periodo l'Europa settentrionale era ricoperta da foreste sia tropicali, sia temperate. I reperti fossili dimostrano che molti insetti di tali foreste assomigliavano agli insetti degli analoghi ambienti dell'Europa e dell'Asia tropicale del periodo attuale. In particolare le specie estinte del genere *Oecophylla* sono più affini all'attuale specie asiatica che non a tutte le altre specie di formiche tessitrici che vivono attualmente in Europa.

Circa 15 anni fa fu compiuta la prima osservazione relativa all'organizzazione sociale dell'*Oecophylla* estinta. Nel 1963 Mary Leakey, in cerca di fossili sull'Isola Mfwangano nel Lago Vittoria, in Kenya, riuscì a scoprire un raggruppamento di formiche fossili, 366 minuscole formiche cristallizzate riunite in un singolo punto. Evidentemente tali insetti vivevano in un nido di foglie che era caduto in uno stagno d'acqua dolce, dove il nido e i suoi abitanti erano stati rapidamente ricoperti da sedimenti. In queste condizioni si sono conservati numerosi particolari superficiali. Uno di noi (Wilson) ha identificato il raggruppamento come una porzione di una colonia di formiche tessitrici estinte. È la prima e per ora l'unica società d'insetti scoperta allo stato fossile. Il raggruppamento comprende grappoli di larve e pupe. Alcune formiche fossili sono ancora attaccate a frammenti di foglie. Poiché nel blocco c'è una piccola popolazione, anziché i soliti esemplari fossili isolati, è stato possibile compiere uno studio statistico del sistema di caste nella specie fossile. Le caratteristiche anatomiche di queste formiche fossili e l'abbondanza relativa delle due caste di operaie si sono rivelate del tutto simili a quelle che, tra tutte le formiche attuali, sono tipiche delle specie del genere *Oecophylla*.

L'elemento più caratteristico del sistema sociale delle attuali formiche tessitrici è costituito dalle operaie minori, cioè da quella casta di formiche più piccole e meno numerose che si sono specializzate nella cura delle uova e delle piccole larve. Nella maggior parte delle specie di formiche che hanno più d'una casta sono le operaie maggiori a risultare meno numerose e devianti dal punto di vista anatomico. La popolazione fossile scoperta da Mary Leakey era in un deposito del Miocene inferiore e ha perciò un'età di almeno 15 milioni d'anni. Le particolarità dell'anatomia e della distribuzione delle dimensioni, che si riscontrano tanto nelle formiche del Miocene, quanto in quelle attuali, fanno pensare che la particolare divisione del lavoro tra le specie viventi di *Oecophylla* risalga a tempi assai remoti.

Questo documento fossile fa pensare che l'organizzazione sociale avanzata conferisca una stabilità evolutiva alle società degli insetti. Questa stabilità presenta vantaggi e svantaggi. Da un lato, come i nostri studi hanno dimostrato, ogni operaia esibisce non più di 50 atti comportamentali distinti, la maggior parte dei quali sono a fini di comunicazione. La conseguenza di questo orientamento quasi esclusivamente sociale è che la colonia delle formiche tessitrici è un'unità che lavora con estrema efficienza. D'altra parte però l'efficienza della colonia viene ottenuta con una rigida programmazione delle componenti relativamente semplici di comportamento individuale, che garantisce un modello complesso ma stereotipato di cooperazione durante le attività di gruppo. È chiaro che la colonia può prosperare solo a spese di ogni espressione di azione indipendente da parte dell'individuo. In altre parole le formiche tessitrici, così come del resto le api, sembrano aver raggiunto l'estremo limite d'una radiazione adattativa tra gli insetti sociali.

# Un sito rurale celtico nella Britannia meridionale

*Dal VI secolo a.C. fino ai decenni seguenti alla conquista romana della Britannia, tre successivi insediamenti di agricoltori e pastori celti occuparono un'area circondata da un fossato nel Dorset*

di Geoffrey Wainwright

Negli ultimi secoli prima dell'era cristiana, la maggior parte dell'Europa compresa fra il Danubio e l'Atlantico era controllata da tribù celtiche. Una volta i celti riuscirono addirittura a saccheggiare Roma. Nello stesso periodo ondate successive di celti si spinsero a popolare la maggior parte della Britannia. Le tribù quivi immigrate dominarono l'isola fino alla conquista romana e, anche in seguito, dopo essersi romanizzate, conservarono gran parte della loro autorità sino all'epoca degli anglo-sassoni. Ma anche gli studiosi di storia antica sanno ben poco sulla vita rurale nella Britannia celtica. Gran parte di questo stato di ignoranza va addebitata agli stessi archeologi britannici.

Fino a cinque anni or sono l'unico insediamento rurale celtico di una certa estensione in Gran Bretagna che fosse stato scavato sistematicamente e che avesse fornito informazioni socioeconomiche significative sulla Britannia preromana era Little Woodbury, un insediamento recintato nel Wiltshire, presso Salisbury. Gerhardt Bersu vi diresse i lavori nel 1938 e 1939 per incarico della Prehistoric Society, ma le ricerche furono interrotte quando solo un terzo dell'area era stato riportato in luce. Benché parziali, gli scavi consentirono nondimeno deduzioni utili circa la cultura, la società e l'economia della Britannia celtica, ma l'incompletezza del lavoro impedì di raggiungere conclusioni sicure su vari punti.

Successive ricerche sul campo e l'osservazione aerea nell'Inghilterra meridionale rivelarono un certo numero di insediamenti paragonabili a Little Woodbury, compreso un sito agricolo-pastorale celtico di dimensioni abbastanza grandi nel Dorset, 25 chilometri circa a nord del porto di Bournemouth, sulla Manica. I terrapieni originali erano scomparsi ormai da molto tempo, cancellati dall'aratro, ma in fotografie aeree scattate da J.K.S. St. Joseph risultarono visibili i contorni di un'area recintata di 1,2 ettari. Per estensione e configurazione l'area era molto simile al sito di Little Woodbury. Questo nuovo sito offrì perciò una opportunità di estendere le ricerche di Bersu, questa volta per mezzo di uno scavo dell'intera superficie. I lavori nell'*enclosure* di Gussage, così chiamata dal vicino paese di Gussage All Saints, furono iniziati nel 1972 per conto del British Department of the Environment. A causa della continua erosione dovuta all'aratura, lo scavo fu classificato come una operazione di salvataggio archeologico.

Il sito di Gussage si trova nel cuore di un territorio occupato nel I secolo a.C. da una tribù celtica, quella dei durotrigi. La tribù controllava il Dorset e parti delle contee adiacenti del Wilts, del Somerset e del Devon. Il territorio dei durotrigi e quello delle tribù celtiche vicine è contrassegnato da fortificazioni sulla cima di colline, come Hod Hill e una grande *enclosure* di 18 ettari nota come Maiden Castle. Le difese collinari furono occupate per circa sei secoli da vari gruppi celtici prima che questa parte della Britannia venisse pacificata nel 43-44 d.C. dalla Seconda Legione (Augusta), comandata dal futuro imperatore Vespasiano.

I nostri scavi dimostrarono che il sito agricolo-pastorale di Gussage era stato popolato pressappoco per il medesimo periodo di tempo. Il sito era stato occupato per la prima volta da una popolazione stabile attorno al 550 a.C. e fu abbandonato infine fra il 50 e il 75 d.C., non molto tempo dopo l'arrivo dei romani. Da un punto di vista strutturale, lo sviluppo dell'insediamento ebbe luogo in tre fasi distinte.

Il primo insediamento, fondato all'inizio dell'Età del ferro britannica, durò per due secoli e mezzo, dal 550 a.C. circa fino al 300 a.C. circa. Attorno a esso correva un fossato di poco più di un metro di larghezza e di meno di un metro di profondità; il terriccio asportato dal fossato fu utilizzato per erigere un piccolo argine sul bordo esterno del fossato stesso. L'ingresso principale nell'area recintata era a est. Due fossati sussidiari «ad antenna» servivano a incanalare il traffico in entrata verso una strada che attraversava il fossato principale e terminava in corrispondenza di una porta in legno. La combinazione dei fossati ad antenna e della porta d'ingresso era ispirata evidentemente a ragioni difensive.

Le strutture costruite all'interno del primo insediamento di Gussage ci sono note solo dai fori in cui erano piantati i pali che un tempo sostenevano gli edifici in legno. Analogamente, le attività della comunità agricolo-pastorale ci sono suggerite solo dai contenuti di varie fosse utilizzate per la discarica di rifiuti che troviamo associate alle prime strutture e dagli avanzi e scarti di attività artigianali rinvenuti in talune depressioni che dovevano ospitare fucine o laboratori all'interno dell'area. La maggior parte delle fosse e delle depressioni si trovavano a qualche distanza dal fossato perimetrale. L'ordine della loro distribuzione è tipico: esse sono adiacenti a gruppi di fori per pali, molti dei quali formano dei quadrati, con un foro in ogni angolo. I lati dei quadrati hanno una lunghezza variabile fra 1,7 e 2,5 metri.

I quadrati così formati costituiscono,

**Vista aerea della comunità agricolo-pastorale celtica riportata in luce nei pressi del paese di Gussage All Saints, nel Dorset (Inghilterra); la fotografia illustra, attraverso la successione di fossati, di fori per pali e di buche per i rifiuti, sei secoli di occupazione del sito. Per mezzo di uno scavo, condotto sull'intera superficie, il gruppo di lavoro diretto dall'autore ha svolto un'importante azione di recupero, salvando i resti che esso custodiva da altri eventuali danni.**

Le macchie e le linee scure che traspaiono attraverso le colture, nei campi coltivati visibili in questa fotografia aerea obliqua, condussero all'individuazione dell'insediamento di Gussage. Il nord è in alto; si vedono distintamente la maggior parte del fossato perimetrale ed entrambi i fossati «ad antenna», come pure il fossato anulare scavato verso la fine dell'ultima fase. La maggior parte delle macchie scure evidenziano fosse nelle quali venivano scaricati i rifiuti. Le macchie nella vegetazione derivano dal fatto che le piante che crescono in un suolo irregolare spingono le radici più in profondità, hanno un migliore accesso alle acque sotterranee e acquistano un colore verde più scuro.

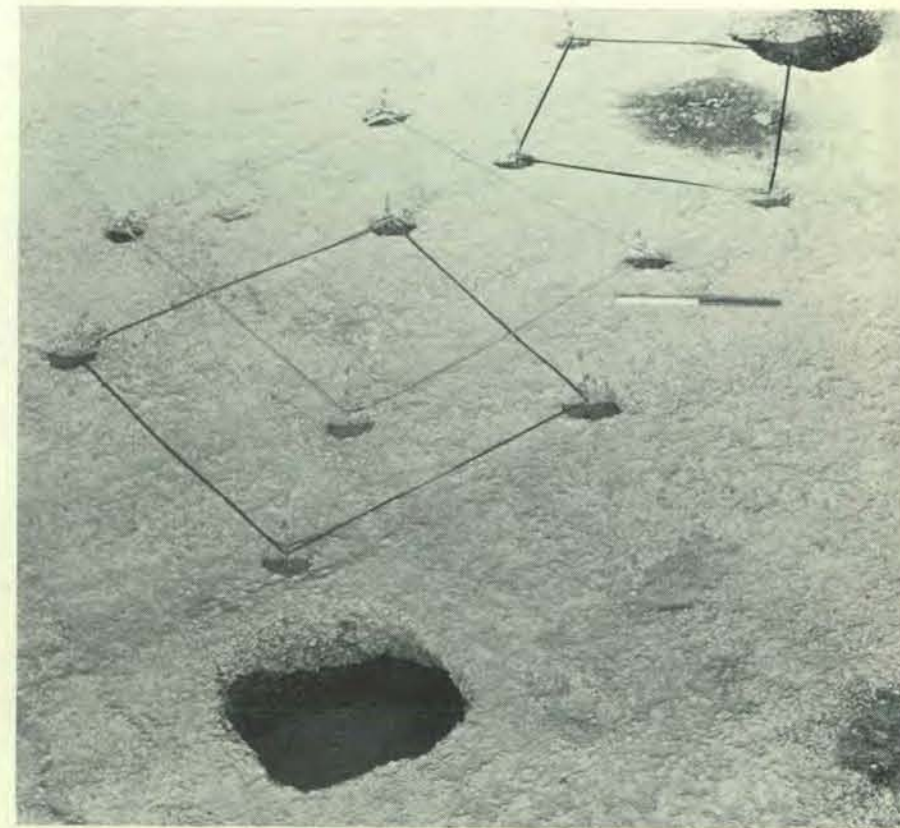
Il fossato anulare dell'ultima fase di Gussage formava una recinzione difensiva all'interno del perimetro principale dell'insediamento. Esso fu scavato a metà del I secolo d.C. All'interno di questo anello una struttura circolare di legno era difesa ulteriormente da un robusto portone di legno. Il fossato circolare si riempì rapidamente di melma e di detriti e il fatto che non si fece alcuno sforzo per mantenerlo pulito e funzionante sembra indicare che le esigenze difensive che ne avevano determinato lo scavo furono probabilmente di breve durata.

per il primo insediamento di Gussage, l'unico resto di quelle che potevano essere abitazioni. Occorre però tener presente che secoli di erosione dovuta all'aratura hanno cancellato ogni prova dell'esistenza di strutture di dimensioni minori. Risultano individuabili da 15 a 20 quadrati, concentrati in un'area di circa 50 metri per 40 al centro dell'area recintata. Le fosse associate a essi sono in numero di 125; ogni fossa in media ha un volume di meno di due metri cubi, cosicché lo spazio totale disponibile nelle fosse della prima fase a Gussage è di 229 metri cubi. Alcune di queste fosse potrebbero essere state usate per conservarvi del grano, ma alla fine furono riempite tutte di rifiuti e dei resti sia di animali sia di esseri umani.

La seconda fase a Gussage occupò un periodo compreso pressappoco fra il 300 e il 100 a.C.. Taluni indizi suggeriscono l'esistenza di una continuità fra le due fasi. Per esempio, benché l'ingresso all'area recintata venisse ricostruito, la sua ubicazione non fu cambiata e il fossato principale, pur essendo stato ingrandito, continuò a seguire il vecchio tracciato. Inoltre le fosse appartenenti alla seconda fase furono scavate quasi tutte a una certa distanza da quelle della prima fase. Le fosse appartenenti a questo secondo periodo sono però decisamente più grandi di quelle anteriori. Pur essendo solo 69, hanno complessivamente una capacità di circa 210 metri cubi.

Più scarsa è la documentazione architettonica relativa alla seconda fase. L'unica struttura sicuramente residenziale era a pianta circolare. Come attestano le tracce di un condotto di scolo attorno a essa, doveva avere un diametro di circa nove metri. Deboli tracce di un secondo condotto del genere suggeriscono l'esistenza di un'altra struttura circolare analoga durante questa fase, ma non c'è alcun segno di fori per i pali. Quanto al fossato perimetrale attorno a questo secondo insediamento, la sua larghezza fu portata a più di due metri e la sua profondità a circa un metro e mezzo. L'ingresso nella zona recintata fu ristrutturato, compreso lo scavo di un secondo paio di fossati ad antenna. La larghezza del passaggio fu ridotta a cinque metri e fu introdotto un sistema più elaborato con porta in legno inserita in una palizzata con fossato. I fori per i pali di sostegno sono abbastanza grandi da suggerire che la porta poteva essere sovrastata da una torre, ma l'ingresso potrebbe non aver superato i due metri e mezzo di larghezza.

La terza e ultima fase a Gussage durò dal 100 a.C. circa fino all'abbandono dell'insediamento, che avvenne attorno all'80 d.C. Anche qui un certo grado di continuità fra fasi successive è suggerito dal fatto che tanto l'ingresso quanto il fossato perimetrale conservarono l'ubicazione anteriore. Le 180 fosse scavate durante la fase finale a Gussage furono però distribuite indipendentemente dalla posizione delle buche anteriori. Solo leg-

Un mucchio di ossa accumulate alla rinfusa fu riportato in luce nella parte più alta della fossa numero 293, una delle fosse associate alla fase finale di Gussage. Sono state individuate le ossa spezzate di arti, mandibole di bovini e i crani di un bue e di un cavallo. Gli abitanti di Gussage allevavano bovini, pecore e capre, mentre si procuravano i cavalli catturandoli.

Alcune serie di quattro fori per pali disposti in quadrato, risalenti alla prima fase di Gussage, sono qui marcate con funicelle; le immagini così ottenute suggeriscono che i pali piantati in tali fori dovevano formare i sostegni principali di strutture residenziali. La lunghezza dei lati dei quadrati varia da 1,7 a 2,5 metri. È improbabile che i pali piantati in questi fori fossero pilastri d'angolo, perché edifici così piccoli avrebbero offerto un rifugio estremamente disagevole.

0 10 20 30 40 50
METRI
NORD
LIMITE DEGLI SCAVI

germente più piccole delle fosse della seconda fase, le nuove fosse hanno un volume complessivo che si approssima a 500 metri cubi.

Durante la fase finale all'interno del fossato perimetrale furono costruite tre aree recintate supplementari. Le due più importanti sono rispettivamente un'area trapezoidale che potrebbe essere stata utilizzata come recinto per il bestiame e un consistente terrapieno circolare che era destinato probabilmente a fornire maggior sicurezza alle strutture situate nel suo interno. Ma quale che fosse la ragione che ispirò lo scavo di questo fossato circolare, esso fu costruito piuttosto tardi nella storia dell'insediamento, ossia attorno alla metà del I secolo d.C. Il fossato si ostruì peraltro rapidamente senza che nessuno si prendesse la briga di mantenerlo in ordine; ciò, quindi, suggerisce l'idea che il bisogno di una tale opera difensiva sussidiaria sia venuto meno molto presto.

Finora ho fornito le date delle fasi successive dell'insediamento di Gussage senza citare alcuna prova a sostegno. Sul sito furono trovati un certo numero di resti di carbone di legna, idonei per la datazione al radiocarbonio. Nove di questi campioni furono inviati a V.R. Switsur, del Radiocarbon Dating Research Laboratory dell'Università di Cambridge. Tre resti di carbone provenienti dalla prima fase fornirono datazioni comprese fra il 790 e il 420 a.C. Due campioni provenienti da contesti della seconda fase dettero misurazioni comprese fra il 410 e il 140 a.C., e altri due resti, della terza fase, furono datati al periodo 50 a.C. - 150 d.C.

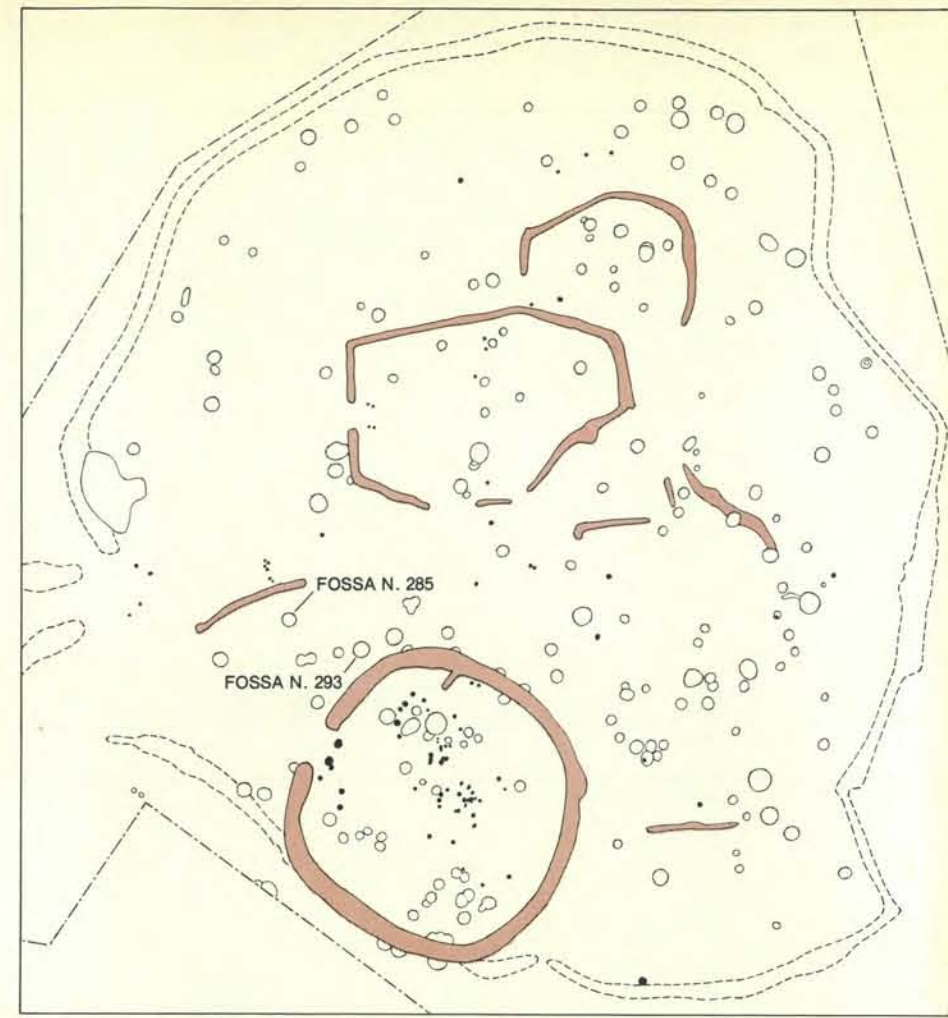
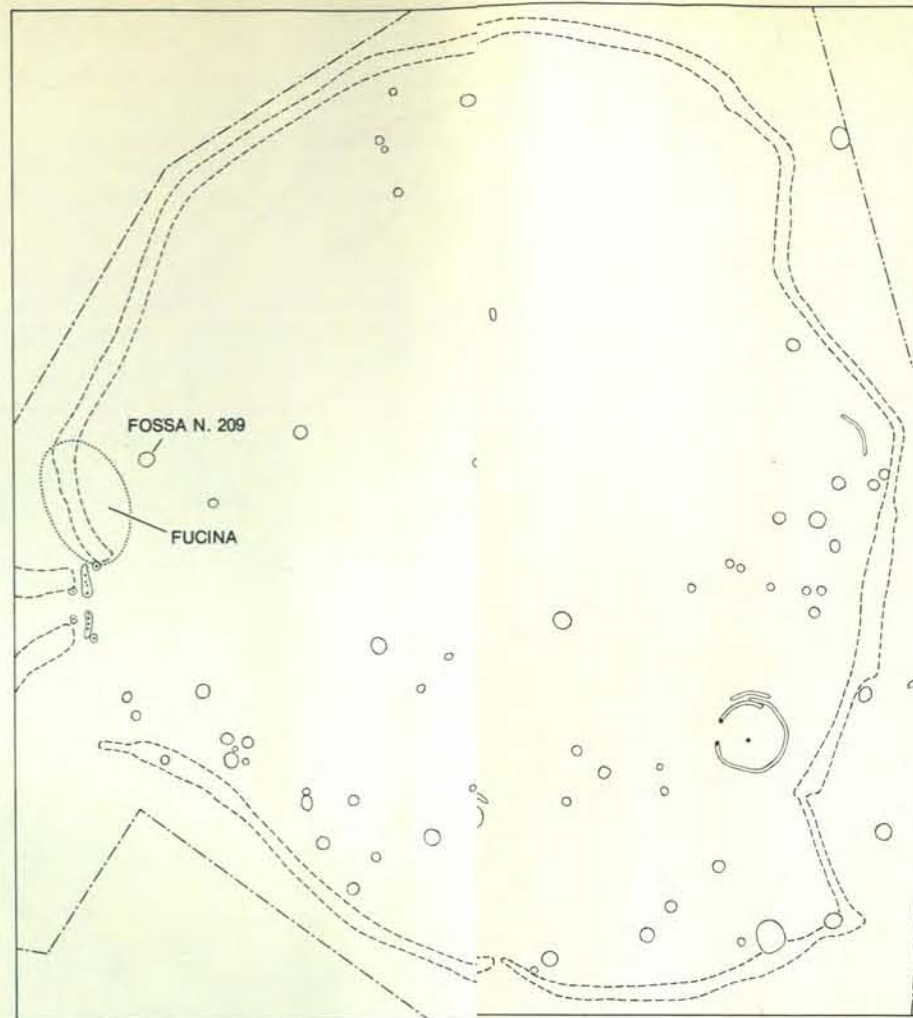
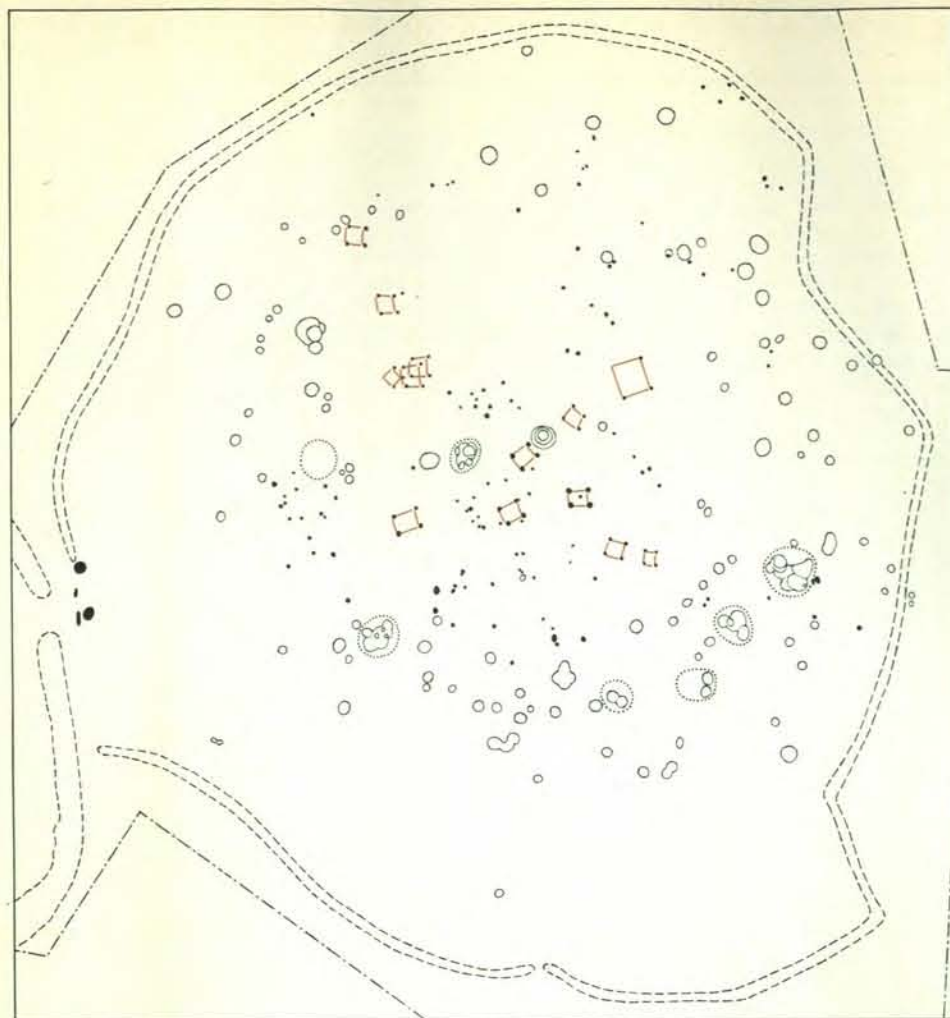
Tali intervalli cronologici sono molto ampi e la vera età può trovarsi in qualsiasi punto fra i due estremi. Per l'abbandono del sito agricolo-pastorale di Gussage sono disponibili però dati un po' più precisi. Un tipo di ceramica rossa molto ammirata dai romani e classificata in modo piuttosto vago come «samia» fu soggetta a mutamenti stilistici così costanti che vari tipi di ceramica samia possono essere considerati con un buon grado di sicurezza come indicativi di vari periodi imperiali. Cinque fra le fosse riempite di rifiuti a Gussage fornirono cocci di ceramica samia associati ai periodi di regno degli imperatori romani Claudio (41-54 d.C.) e Flavio (69-79 d.C.). La presenza di questi cocci dimostra la persistenza dell'insediamento oltre la metà del I secolo d.C., e probabilmente sino all'ultimo quarto del secolo. Quanto alla più antica fra le datazioni al radiocarbonio, indica la presenza di un insediamento iniziale a Gussage qualche tempo prima della metà del primo millennio a.C.

Se si fa eccezione per la ceramica, i manufatti rinvenuti nelle fosse dei rifiuti e nelle depressioni usate come laboratori artigianali a Gussage non sono particolarmente numerosi. Tuttavia, congiuntamente ai resti organici a essi associati, ci forniscono indizi preziosi di natura tanto sociale, quanto economica sulla vita dei celti dell'Età del ferro nella Britannia rurale. Per esempio, nel campo della tecnica, la macina a sella, uno strumento abbastanza primitivo per macinare il grano, era ancora in uso all'inizio della prima fase a Gussage. Non molto tempo dopo la fondazione dell'insediamento, però, fu introdotto uno strumento molto più efficace, la macina girevole. Per fare un altro esempio, i pettini per cardatura in osso sono indicativi di un qualche grado di sofisticazione nella manifattura di tessuti. Ora, tali pettini sono praticamente assenti dai resti dell'insediamento della prima fase, mentre sono relativamente numerosi nella seconda fase.

Un altro progresso tecnico è attestato dalla ceramica di Gussage. Le ceramiche della prima e della seconda fase sono ceramiche convenzionali, foggiate a mano, tipiche della Britannia meridionale di questo periodo, per esempio vasi globulari, belle ciotole rivestite di ematite, pentolini con pareti verticali e orci con pesante piedestallo. Nella fase finale a Gussage, in coincidenza con l'introduzione della ruota da vasaio, gli stili della ceramica conobbero una rapida evoluzione. Fra il vasellame foggiato con la ruota possiamo citare boccali, grandi brocche con manici e ciotole con anelli al piede, per menzionare solo alcuni tipi.

Gli oggetti di uso personale non sono numerosi in nessuna delle tre fasi. Tipici di quelli rinvenuti nei depositi di rifiuti della prima fase a Gussage sono spilli di ferro con testa ad anello, fermagli di ferro e bronzo (compreso un fermaglio di bronzo tipico del periodo celtico quale è noto al di là della Manica, come a La Tène 1-*a*), braccialetti in bronzo e ardesia, perle di vetro e cavicchi d'osso. La seconda fase è più povera anche di oggetti comuni come questi. Nella fase finale appaiono in quantità crescente fermagli, pendenti d'osso e una varietà di articoli per toeletta.

Fra i manufatti più familiari vi sono punte di ferro per vomeri e coltelli di ferro con tipiche punte sdoppiate. Una bilancia a bracci uguali in bronzo attesta la pratica della pesatura, così come un peso di ferro per una stadera. Nei rifiuti di tutt'e tre le fasi si trovano oggetti che possono essere considerati indicativi dello status sociale, come frammenti di fodero, che testimoniano l'uso di portare la spada. Può darsi però che si tratti di

**La pianta qui a sinistra presenta le caratteristiche principali delle tre fasi successive dell'insediamento di Gussage. L'orientamento della piantina è lo stesso della fotografia di pagina 94; l'est si trova a sinistra. L'area recintata, che ha un diametro di circa 110 metri da est a ovest e di oltre 120 metri da nord a sud, copre 1,2 ettari. I fossati ad antenna sul lato a est servivano a incanalare il traffico verso la porta d'ingresso. Il colore indica una fossa di data sconosciuta.**

**Le tre fasi d'occupazione individuate a Gussage sono datate rispettivamente ai periodi 550-300 a.C. (*a sinistra*), 300-100 a.C. (*al centro*) e 100 a.C. - 80 d.C. (*a destra*). (Gli estremi cronologici dei tre periodi sono ovviamente approssimativi.) Il fossato che circonda il primo insediamento era a forma di *V*, aveva una profondità di 80 centimetri e una larghezza di 1,2 metri. Due grandi fori per pali rivelano che la porta d'ingresso si trovava sul lato orientale dell'area cintata; i fossati ad antenna all'esterno dell'ingresso erano larghi tre metri. Molti fori per pali all'interno dell'area formano quadrati. Può darsi che contenessero i principali pilastri delle strutture residenziali. Tredici di essi appaiono in colore nella piantina. Sette laboratori artigianali (*linee punteggiate*) sono associati alla prima fase d'insediamento a Gussage. Le fosse per la discarica dei rifiuti di questa fase avevano una capacità complessiva di 229 metri cubi. Che fra la prima e la seconda fase non ci sia stata soluzione di continuità è suggerito dal fatto che l'ingresso, pur essendo stato ricostruito, non ricevette però un'ubicazione diversa e che il fossato perimetrale, pur essendo stato scavato più in profondità e portato a una larghezza doppia, conservò il medesimo tracciato. Delle abitazioni interne all'area delimitata dal fossato e appartenenti alla seconda fase, solo alcuni fori di pali associati a una struttura circolare sono sfuggiti alla distruzione a opera delle attività agricole. La capacità complessiva delle 69 fosse scavate durante la seconda fase è di circa 210 metri cubi. La fossa 209, adiacente a una grande depressione che ospitava una fucina, conteneva i rifiuti della fusione di accessori in bronzo per cavalli e per carri del II secolo a.C. I caratteri principali della fase finale di Gussage sono i resti di recinti e un fossato circolare con molti fori per pali. Nella fossa 285 sono stati rinvenuti i resti di un uomo che presentavano segni di morte violenta.**

accumuli di semplici rottami metallici; tanto la metallurgia del ferro quanto quella del bronzo furono praticate nell'insediamento di Gussage per tutta la durata della sua storia.

Semi carbonizzati e indizi analoghi della pratica dell'agricoltura a Gussage indicano che la popolazione dei vari insediamenti succedutivisi coltivava frumento e orzo, probabilmente il primo in quantità maggiori rispetto al secondo. Il frumento era di una varietà oggi rara: la spelta (*Triticum spelta*). Sin dalla prima fase degli insediamenti di Gussage fu coltivata anche l'avena, e così pure piselli e fagioli. In effetti la crescente abbondanza di legumi nel corso dei secoli suggerisce che essi dovettero acquistare sempre maggiore importanza come generi alimentari. Tracce dei campi in cui queste piante venivano coltivate sono ancora visibili nelle zone a nord, a ovest e a sud dell'area cintata. Il raccolto veniva evidentemente fatto essiccare, tutto o in parte, prima di essere trebbiato e immagazzinato. A questo scopo venivano usati forni di argilla, molti dei quali decorati; i resti frammentari di questi forni per l'essiccazione sono particolarmente numerosi nella fase finale. Alcune delle fosse scavate a Gussage potrebbero essere servite per immagazzinarvi la granella una volta che era stata spulata.

Sulla base dei 15 500 esemplari di ossa di animali ritrovati a Gussage è possibile ricostruire con qualche certezza la natura dell'economia pastorale dell'insediamento. Fra le specie rappresentate in questo ricco campionario di ossa vi sono cavalli, bovini, pecore, capre, maiali, cani e gatti. Le ossa rappresentano in modo più o meno uniforme tutte le parti del corpo delle sette specie. Tutti i gruppi di età, dallo stadio fetale, al periodo immediatamente successivo alla nascita, all'età avanzata, sono rappresentati in proporzioni tali, con l'eccezione dei cavalli, da suggerire che ci troviamo di fronte a popolazioni «naturali» dinamiche.

I reperti animali furono analizzati da R.A. Harcourt. Egli trovò un picco di mortalità nel periodo compreso fra i quattro e i sei anni di età; questo risultato conservò la sua validità per l'intera storia dell'insediamento. Fra i bovini tale tipo di mortalità indica l'eliminazione sistematica degli esemplari giovani che non venivano utilizzati come animali da tiro; essi venivano uccisi presumibilmente per consumarne la carne. Venivano macellate anche le vacche, forse nei casi di sterilità o di scarsa produzione di latte o di altri difetti.

Fra gli ovini il tipo di mortalità riscontrato indica un tasso del 5 per cento di decessi per cause naturali subito dopo la nascita e un'eliminazione selettiva di un 16-26 per cento degli agnelli di cui non c'era bisogno per il mantenimento di un gregge. È questo il tipo di comportamento che ci si attenderebbe se il gregge fosse conservato primariamente per avere una produzione costante di lana e latte, con la carne come prodotto secondario.

La gamma di età documentata dalle numerose ossa di cavallo pone questi animali in una posizione a sé rispetto a tutti gli altri animali domestici a Gussage; nelle ossa rinvenute non sono rappresentati né puledri appena nati né esemplari di un anno di età. I reperti consentono una sola interpretazione: non era praticato l'allevamento dei cavalli. Doveva essere invece abituale l'uso di catturare periodicamente puledri selvatici e dalla mandria dovevano venire selezionati cavalli maturi di tre anni per addestrarli ai finimenti. La maggior parte dei cavalli di Gussage rientra nella gamma di grandezza della razza propria dell'Età del ferro inglese: altezza alla spalla compresa fra 110 e 145 centimetri. (Un cavallo da sella attuale è alto alla spalla circa 180 centimetri.)

Una parte degli abitanti del sito agricolo-pastorale di Gussage doveva essere occupata in industrie «domestiche». Per esempio, un certo numero dei recipienti di ceramica della fase finale ha una base perforata, cosa che fa pensare che venissero usati per l'eliminazione del siero del latte nella produzione di formaggio. È probabile che le mandrie e i greggi di Gussage fornissero latte in eccesso rispetto alle esigenze della popolazione e una tale eccedenza poteva essere convertita utilmente in formaggio. Non pare quindi azzardato supporre che a Gussage venisse prodotto del formaggio. Per fare un altro esempio, fra i manufatti rinvenuti a Gussage vi sono coltelli d'osso che possono essere interpretati come strumenti per scuoiare il bestiame. Forse le regolari uccisioni di bovini e ovini fornivano la materia prima per una industria delle pelli. Infine l'esistenza della tessitura a livello domestico è attestata dall'esistenza di pettini in osso per la cardatura e di numerosi pesi da telaio e dischi per fusi.

I resti di scheletri umani rinvenuti a Gussage rappresentano un minimo di 53 individui: 15 adulti e 38 bambini piccoli. Soltanto uno scheletro fu trovato in un contesto di oggetti della prima fase; era quello di un bambino piccolo sepolto in una fossa scavata appositamente. La maggior parte dei resti umani, rappresentanti 14 adulti e 31 bambini piccoli, fu trovata in contesti della terza fase. Erano stati interrati disordinatamente in fosse per rifiuti e in fossati con scarse o addirittura inesistenti indicazioni di sepolture formali, come disposizione uniforme della salma, orientamento costante del corpo o presenza di arredi funerari.

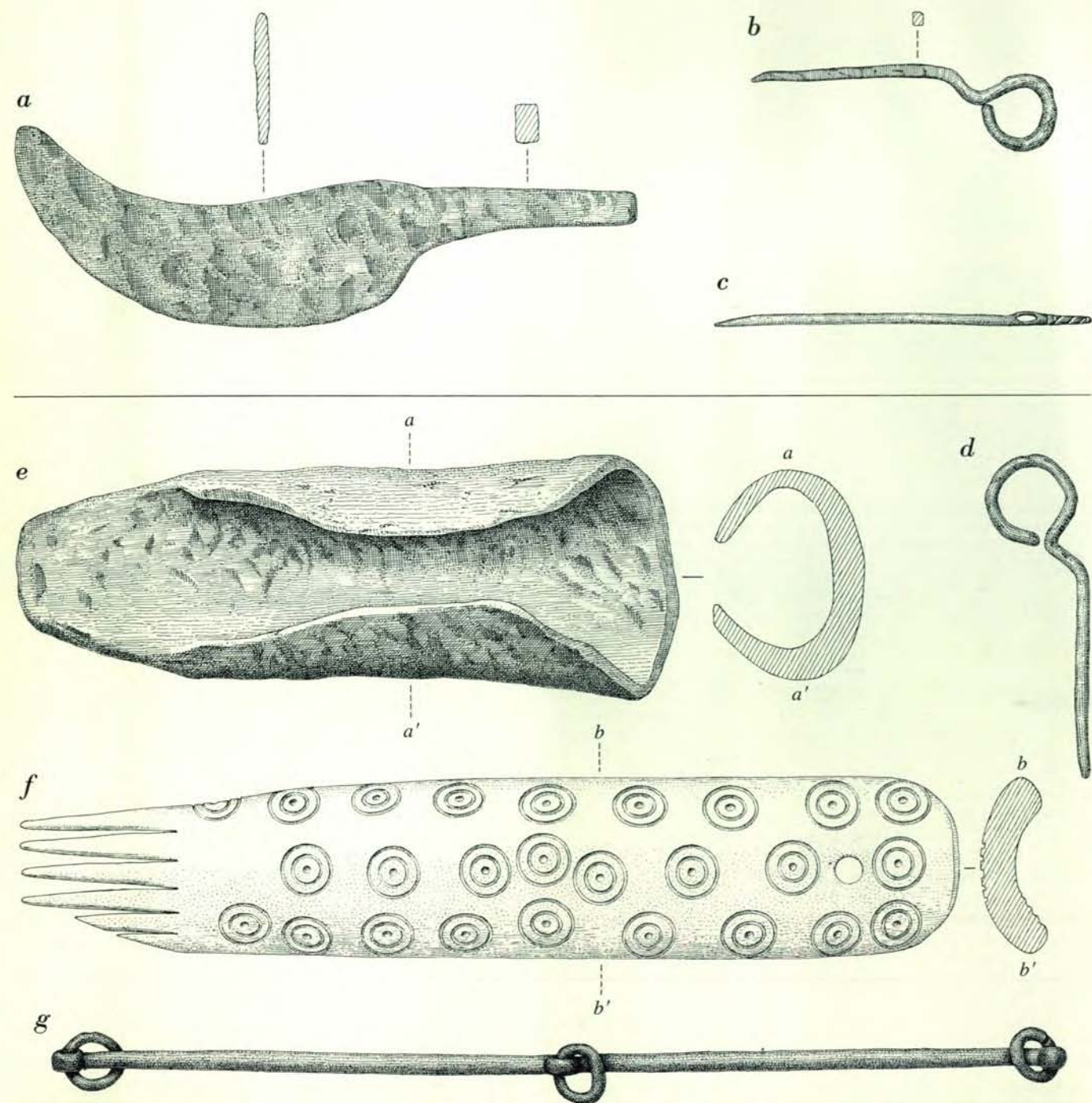
Può darsi che, al di fuori dei confini dell'insediamento, esistessero cimiteri regolari, ma finora non ne è stato individuato alcuno. Frammenti sparsi di ossa umane sono però relativamente numerosi a partire dalla seconda fase dell'insediamento, quando furono interrati, senza tante formalità, un adulto e sei bambini piccoli. Questi reperti suggeriscono la possibilità che esistessero usi funerari come un'esposizione preliminare del corpo e che poi ci si liberasse senza tante cerimonie delle ossa.

Carole Keepax, dell'Ancient Monuments Laboratory, ha analizzato i resti umani rinvenuti a Gussage. Ella ha riscontrato un'incidenza abbastanza elevata di malattie degenerative delle ossa (osteoartrite) e un tasso molto alto di carie dentaria. La diffusione di quest'ultima infermità potrebbe essere stata causata da un fattore dietetico non ancora identificato.

Abbastanza alto risultò anche il numero delle fratture guarite. Di particolare interesse è lo scheletro di un robusto giovane di sesso maschile trovato in prossimità della parte superiore della fossa numero 285, appartenente alla fase finale di Gussage. Il cranio e le ossa del braccio sinistro rivelano le tracce di colpi inflitti con uno strumento acuminato. I colpi furono inferti subito prima della morte o forse anche subito dopo. Le ferite fanno pensare a una lotta faccia a faccia cosicché è probabile che il giovane sia perito di morte violenta.

La scoperta più sorprendente fatta a Gussage è costituita dai rottami provenienti da una fucina per la lavorazione del bronzo, in un deposito di rifiuti risalente alla seconda fase: la fossa numero 209. Lo scavo di questa fossa fu eseguito con diligenza da Karem Stanley e i reperti relativi sono stati analizzati da un'autorità nel campo della metallurgia dei celti, Mansel C. Spratling, del British Museum. Il fabbro che fondò questa fucina a Gussage, in epoca imprecisata durante il II secolo a.C., produsse una cinquantina di accessori in bronzo per finimenti destinati ai pony e ai carri, usando il metodo di fusione a cera perduta. Spratling stima che il fabbro abbia

a b c

a e a a' a' d b f b b' b' g

**Fra i manufatti celtici rinvenuti nelle fosse di rifiuti a Gussage vi sono un coltello di ferro (*a*), uno spillo di ferro con testa ad anello (*b*) e un ago di bronzo con decorazioni al di sopra della cruna (*c*), tutti appartenenti alla prima fase ossia al periodo compreso tra il 550 e il 300 a.C. Un altro spillo con la testa ad anello, questa volta di bronzo (*d*), una punta di ferro per vomere (*e*), un pettine per cardatura in osso (*f*) e un giogo di bilancia in bronzo (*g*) sono rappresentativi dei manufatti trovati fra i rifiuti della seconda fase tra il 300 e il 100 a.C.**

consumato nel processo di fusione e getto di un totale di 38 chilogrammi di bronzo, oltre a una scorta di carbone di legna per il forno e di argilla per le forme, circa 10 litri di cera d'api.

Forse gli oggetti in bronzo venivano eseguiti semplicemente per essere barattati o venduti. Se a Gussage venivano costruiti anche carri e se i pony di Gussage venivano bardati per trainare tali carri, la comunità doveva aver bisogno di legno stagionato e di attrezzi di ferro per la costruzione dei veicoli e di pelle per i finimenti. Può darsi perciò che la popolazione della comunità agricolo-pastorale comprendesse anche individui in grado di organizzare le operazioni logistiche presupposte da tali attività.

Non è stata trovata alcuna traccia della forgia o degli oggetti in bronzo prodotti. L'area in cui il metallo veniva lavorato doveva essere però vicina alla fossa numero 209, poiché i rottami provenienti dalla forgia non presentano alcun segno di alterazione conseguente a esposizione alle intemperie: evidentemente venivano gettati nella fossa dei rifiuti subito dopo essere stati scartati. I rottami accumulati pesavano più di 200 chilogrammi e occupavano un volume di circa tre metri cubi. La parte principale è costituita da migliaia di frammenti di forme di colata.

Nella fusione a cera perduta il modello in cera, foggiato con grande cura, dell'oggetto in bronzo che dev'essere prodotto, viene racchiuso in una forma o stampo d'argilla che lo avvolge completamente, eccezion fatta per un foro superiore attraverso il quale verrà versato nella forma il metallo fuso. La forma d'argilla viene poi cotta. Durante la cottura la cera fonde ed evapora e l'interno della forma conserva un impronta esatta in negativo del modello di cera volatilizzatosi. Nella forma viene colato del bronzo fuso, che viene poi lasciato raffreddare, dopo di che la forma viene rotta per liberare l'oggetto di bronzo. Erano i frammenti di queste forme che venivano scaricati nella fossa; la loro posizione stratigrafica dimostra che il fabbro si era limitato a fare ogni volta solo poche repliche di un determinato oggetto in bronzo. Egli si liberava poi delle forme spezzate e passava all'esecuzione di altri tipi di oggetti.

I tre principali tipi di oggetti in bronzo prodotti a Gussage erano acciarini per ruote di carro (le biette decorate che si infilano in un incavo all'estremità dell'asse delle ruote per impedir loro di sfilarsi), quattro diversi tipi di anelli per finimenti e morsi per cavallo nella forma di anelli passati l'uno dentro l'altro. Ogni morso richiedeva due getti consecutivi. Gli anelli laterali venivano dapprima fusi separatamente e poi combinati con un modello in cera dell'anello centrale e collocati in una seconda forma. Dopo la cottura della forma e la rimozione della cera, veniva colato nello stampo il bronzo fuso che si sarebbe poi consolidato nell'anello centrale.

Nella fossa numero 209 sono stati rinvenuti, oltre alle forme spezzate, scorie, crogiuoli rotti, pezzi di ganga, molti frammenti di tubi per lo sfogo delle fornaci di argilla, vari utensili di fucina fatti di ferro e d'osso, mucchi di argilla per forme non utilizzata e un unico lingotto non usato di bronzo allo stagno il cui peso è quasi esattamente un decimo di una libbra celtica «pesante» (638 grammi). Complessivamente, i rifiuti dell'attività metallurgica a Gussage non temono confronti per quantità, varietà e stato di conservazione con quelli di alcun altro sito preistorico o protostorico dell'intera Europa.

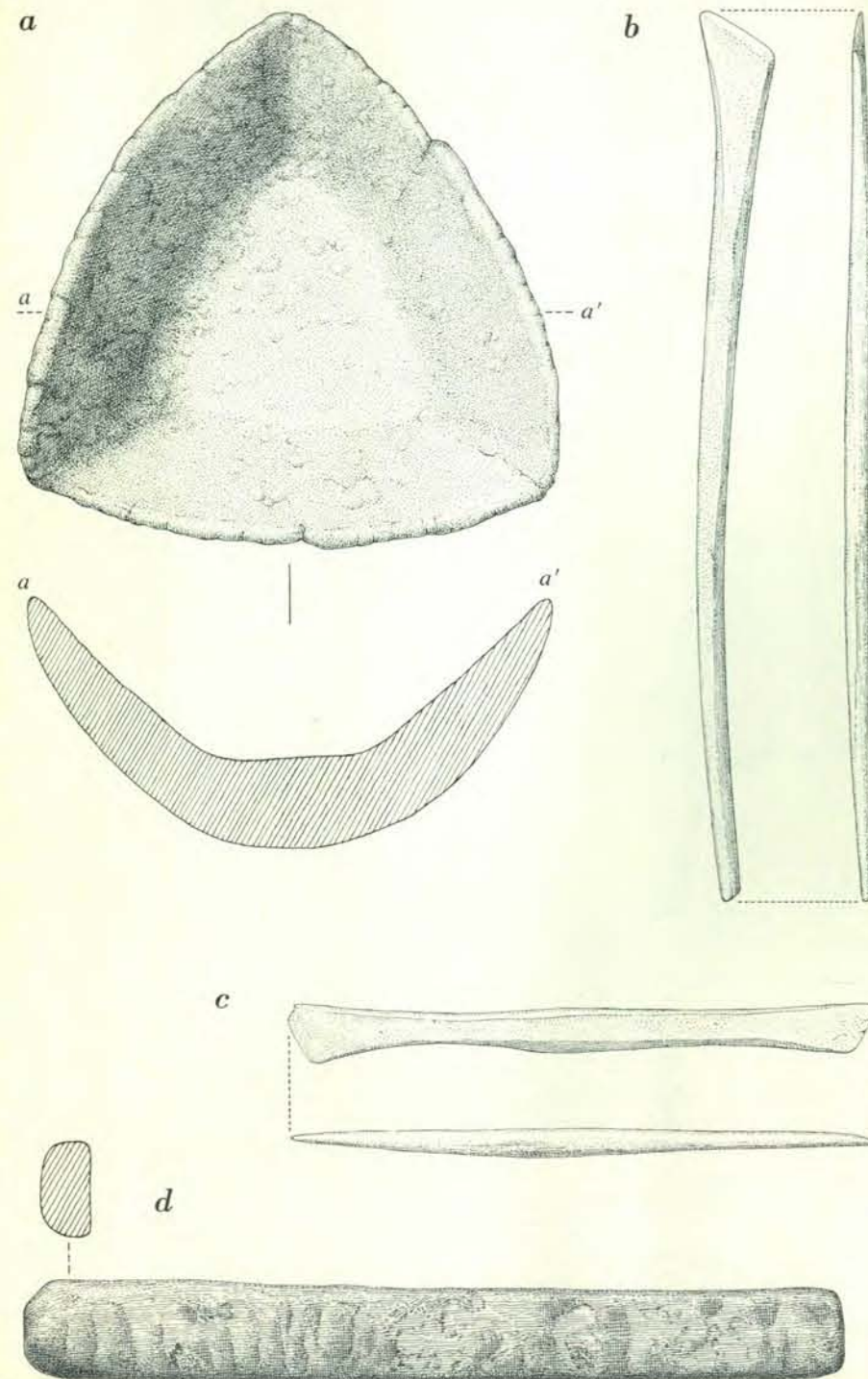
L'analisi dei frammenti di forme da fusione dimostra che il fabbro di Gussage produsse anche fermagli a bottone, fermagli a gancio e lamine metalliche per chiusure. Dei quattro diversi tipi di anelli per finimenti, uno fu prodotto in almeno 11 variazioni. Delle cinquanta e passa serie di bronzi, quelli con una ornamentazione semplice superano in numero quelli più elaborati nel rapporto di 20 a uno. Una tale prevalenza dei modelli più semplici è in netto contrasto col comune concetto delle bardature celtiche decorate per cavalli che è andato formandosi nel corso degli anni sulla base di scoperte avvenute casualmente.

**Fra i rifiuti di una fonderia in cui veniva lavorato il bronzo, provenienti tutti dalla fossa 209, vi sono un crogiolo d'argilla (*a*), utensili d'osso per modellare la cera (*b*, *c*) e un lingotto di bronzo allo stagno (*d*). Questi resti non presentavano segni di esposizione alle intemperie.**



A giudicare dai reperti, venuti in luce in altri sedimenti celtici, la periodica cattura di puledri selvatici, quale risulta dalla distribuzione per età dei resti di cavalli rinvenuti a Gussage, pare sia stata una pratica pressoché generale nella Britannia celtica. È probabile che quest'attività implicasse una cooperazione fra vari insediamenti. Essa potrebbe perciò aver dato luogo a scambi commerciali e sociali fra popolazioni vicine appartenenti alla medesima tribù. L'esistenza di scambi commerciali viene suggerita anche da altri reperti di carattere zoologico. Si ritiene che, fra i bovini domestici con corna, sia sorta in seguito a una mutazione una varietà di bovini senza corna che ebbe una certa diffusione; la prole di tali bovini mutanti non aveva ovviamente corna. A Gussage e altrove nel territorio dei durotrigi assieme a crani di bovini con corna sono stati trovati i crani di bovini senza corna. Se gli esemplari privi di corna erano veramente i discendenti di qualche ceppo mutante, la loro compresenza con bovini con corna è una prova a favore della pratica dello scambio di capi di bestiame.

Attestazioni più concrete dell'esistenza di scambi commerciali sono fornite dalle macine per grano. Queste non venivano prodotte a Gussage bensì acquistate presso centri per la lavorazione della pietra che si trovavano ad almeno 12-13 chilometri di distanza. Analogamente, le ceramiche foggiate a mano risalenti ai primi insediamenti di Gussage venivano prodotte con argille locali o da abitanti della comunità stessa o da specialisti itineranti che lavoravano con materiali locali. Le ceramiche foggiate al tornio, risalenti alla fase finale di Gussage, furono prodotte invece in centri specializzati in qualche località della Britannia meridionale.

**Forme di argilla spezzate, risalenti alla seconda fase di Gussage, usate nella produzione di accessori in bronzo per finimenti di cavalli e per parti di carro col procedimento della cera perduta, conservano impronte negative delle repliche di cera evaporate durante la cottura delle forme stesse. In alto è riprodotta una parte di una forma per l'anello laterale di un morso per cavalli formato da tre anelli. In basso, a sinistra, c'è un altro frammento di forma di un anello laterale con una spira decorativa e, a destra, una decorazione triscele per un acciarino da ruota.**

Non è difficile immaginare che produttori di macine e ceramisti di professione potessero prosperare nella Britannia meridionale vendendo i loro prodotti a insediamenti come Gussage. Le comunità agricolo-pastorali potevano offrire in cambio non soltanto lana greggia, miele, cera, bestiame e cereali, ma anche formaggi, pelli conciate e, a Gussage, anche oggetti in metallo. Supponendo che a Gussage fossero effettivamente disponibili per il commercio eccedenze del genere, potrebbe sembrare sorprendente che i nostri scavi non abbiano portato in luce oggetti di un certo valore, come spade, punte di lancia, scudi, elmi, vasellame di metallo prezioso o alcuni dei famosi gioielli celtici. Bisogna però ricordare che i livelli occupati dalla popolazione di Gussage, e particolarmente quelli posteriori, sono stati soggetti a secoli di erosione da parte dell'aratro.

Non sarebbe saggio, in ogni caso, attribuire all'insediamento di Gussage una posizione molto modesta nella società celtica del tempo, anche ammesso che non vi fosse mai esistito alcun oggetto di prestigio. Consideriamo la robusta porta di legno che consentiva o proibiva l'accesso all'insediamento, con i terrapieni che lo fiancheggiavano, i granai della comunità, le sue mandrie e greggi, i cavalli addomesticati, i tessitori, i fabbri e altri artigiani. Anche in società complesse non c'è una correlazione fissa fra status e possesso di articoli di prestigio. La società celtica, nei limiti in cui può essere ricostruita, esibì caratteri così disparati come franchezza e vanità personale, coraggio e vanagloria, magnanimità, irascibilità, passione per i piaceri della tavola e amore per la poesia epica. In un tale ambiente lo status derivava non tanto da ciò che l'individuo possedeva quanto dall'esempio personale che era in grado di dare e dai contributi da lui forniti a una comunità imbevuta di tradizioni. Attributi come questi sono, com'è ovvio, raramente documentabili sulla base di reperti archeologici. Tuttavia non è impossibile immaginare che agli uomini armati di Gussage venisse riservato un posto onorevole nei ranghi quando si riuniva un'assemblea di guerra dei durotrigi e che Gussage potesse mettere a disposizione, oltre che uomini, anche cavalli per tirare i carri. Può anche darsi che il giovane di Gussage perito di morte violenta, di cui abbiamo parlato in precedenza, stesse difendendo un modo di vita celtico ormai in declino in Britannia.

# GIOCHI MATEMATICI

*di Martin Gardner*

## *Congiungendo con linee insiemi di punti si giunge a diversi (e divertenti) cammini*

«Si dimostri che in un gruppo di sei persone, tre di esse o si conoscono o sono completamente estranee una dall'altra.»

- Problema E 1321

«The American Mathematical Monthly» giugno-luglio, 1958

La mia rubrica di questo mese sull'opera di Ramsey nella teoria dei grafi rende onore a «The Journal of Graph Theory», un periodico uscito quest'anno che tratta esclusivamente di uno dei rami in maggior sviluppo della matematica moderna. Il redattorecapo è Frank Harary, eminente studioso di teoria dei grafi e autore del testo sull'argomento che è più ampiamente usato nel mondo. L'editore è John Wiley & Sons.

La teoria dei grafi studia insiemi di punti congiunti da linee. Due articoli del primo numero del nuovo giornale trattano della teoria dei grafi di Ramsey, un argomento che ha stretta connessione con i giochi matematici. Sebbene nel 1930 siano apparsi alcuni articoli del matematico ungherese Paul Erdös e di altri sulla teoria di Ramsey, fino alla fine degli anni cinquanta non iniziò un serio lavoro di ricerca su quelli che oggi vengono chiamati numeri di Ramsey. Diede grande stimolo a questa ricerca il rompicapo apparentemente innocente citato sopra.

È facile trasformare questo rompicapo in un problema riguardante i grafi. Sei punti rappresentano le sei persone. Si congiunga ogni coppia di punti con una linea usando, per esempio, una matita rossa per indicare due persone che si conoscono e una matita blu per due estranei. Il problema ora è dimostrare che, indipendentemente da come vengono colorate le linee, non si può evitare di ottenere o un triangolo rosso (congiungendo tre conoscenti) o un triangolo blu (congiungendo tre estranei).

La teoria di Ramsey, che tratta questo tipo di problemi, prende il nome da uno straordinario matematico dell'Università di Cambridge: Frank Plumpton Ramsey. Ramsey aveva solo 26 anni quando, nel 1930, morì pochi giorni dopo un'operazione all'addome. Ramsey è noto agli economisti per i suoi notevoli contributi alla teoria economica; ai logici per la semplificazione apportata alla teoria ramificata dei tipi di Bertrand Russel (si dice che Ramsey ha ramseificato la teoria ramificata) e per la divisione dei paradossi della logica nelle classi dei paradossi logici e paradossi semantici. I filosofi della scienza lo conoscono per la sua interpretazione della probabilità in termini di credenza e per la sua invenzione della «proposizione di Ramsey»: un espediente simbolico che rende molto più chiara la natura del «linguaggio teorico» della scienza.

Nel 1928 Ramsey lesse alla London Mathematical Society un articolo, ora classico, intitolato *On a Problem in Formal Logic* (ristampato in *The Foundations of Mathematics*, una raccolta postuma di saggi di Ramsey curata da R.B. Braithwaite e attualmente disponibile in brossura; traduzione italiana in *I fondamenti della matematica*, Feltrinelli, Milano, 1964), in questo articolo Ramsey dimostrò un profondo risultato sugli insiemi che è ora noto come teorema di Ramsey. Lo dimostrò prima per gli insiemi infiniti, osservando che questa dimostrazione era più facile di quella successiva per gli insiemi finiti. Questo teorema, come tanti teoremi sugli insiemi, risultò avere una grande varietà di applicazioni inaspettate ai problemi combinatori. È troppo complicato spiegare qui il teorema nella sua completa generalità; per i nostri scopi, sarà sufficiente vedere come esso si applica alla teoria della colorazione dei grafi.

Quando tutte le coppie ottenibili con $n$ punti sono congiunte da linee, il grafo si dice grafo completo di $n$ punti e si indica con $K_n$. Dato che ci interessiamo solo a proprietà topologiche, non importa come vengono situati i punti o disegnate le linee. Nella figura di pagina 110 si può vedere il modo abituale di illustrare grafi completi da due fino a sei punti. Le linee individuano ogni sottoinsieme di $n$ che ha esattamente due elementi.

Supponiamo di colorare in modo arbitrario le linee di un grafo $K_n$ tutte di rosso o tutte di blu o alcune di rosso e le altre di blu. Questa si dice colorazione a due colori di un grafo. La colorazione è una maniera semplice per dividere tutti i sottoinsiemi di due elementi di $n$ in due classi mutualmente esclusive. Analogamente una colorazione a tre colori li divide in tre classi. In generale, una colorazione a $r$ colori divide le coppie di punti in $r$ classi mutuamente esclusive.

«Sottografo» di un grafo completo è qualsiasi tipo di grafo contenuto nel grafo completo nel senso che tutte le linee e i punti del sottografo sono contenuti nel grafo più ampio. È facile vedere che ogni grafo completo è un sottografo di ogni grafo completo di un numero maggiore di punti. Molti grafi semplici hanno nomi particolari: nella figura di pagina 111 se ne possono vedere quattro famiglie (cammini, cicli, stelle e ruote). Si noti che la ruota a quattro punti e un altro modo di disegnare $K_4$ e viene spesso detta tetraedro perché è una proiezione della struttura di un tetraedro.

Si consideri ora il seguente problema per cui sono necessarie sei matite di differenti colori. A ogni colore associamo un tipo di grafo. Per esempio:

1. Rosso: pentagono (ciclo di cinque punti).
2. Arancione: tetraedro.
3. Giallo: stella a sette punti.
4. Verde: cammino di 13 punti.
5. Blu: ruota di otto punti.
6. Porpora: farfallino (due triangoli aventi esattamente un punto in comune).

Poniamo ora una curiosa domanda. Esistono grafi completi che, se vengono colorati in modo arbitrario con sei colori, contengono sicuramente come sottografo almeno uno dei grafi indicati sopra? In altre parole, indipendentemente da come coloriamo uno di questi grafi completi servendoci delle sei matite, siamo sicuri di ottenere o un pentagono rosso o un tetraedro arancione o una stella gialla a sei punti e così via. Il teorema di Ramsey dimostra che, al di là di una certa dimensione finita, tutti i grafi completi hanno questa proprietà. Il grafo più piccolo di questo insieme infinito si dice grafo di Ramsey per quello specifico insieme di sottografi. Il numero dei suoi punti si dice numero di Ramsey per quell'insieme di sottografi.

Da ogni grafo di Ramsey si può ottenere sia un gioco che un rompicapo. Prendiamo come esempio il gioco seguente. Due giocatori fanno a turno a pescare una delle sei matite e a colorare una linea del grafo di Ramsey. Perde la persona che completa per prima la colorazione di uno dei sottografi specificati. Dato che si tratta di un grafo di Ramsey, non può finire in pareggio. Si tratta anzi del più piccolo grafo completo per cui non è possibile pareggiare.

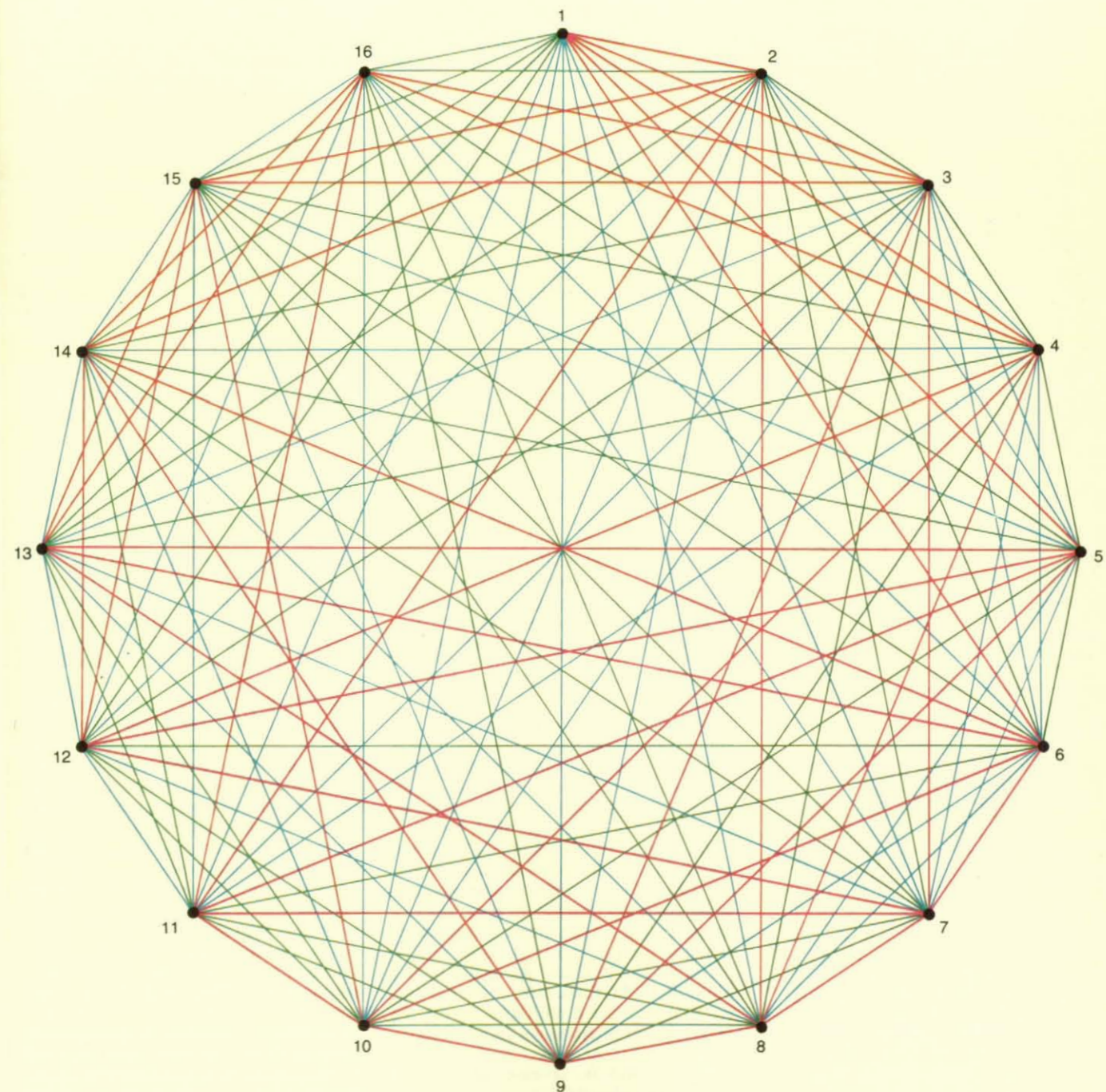
Il relativo rompicapo richiede un grafo completo con un punto in meno del grafo di Ramsey. Questo è ovviamente il più grande grafo completo su cui il gioco può finire in pareggio. Tale grafo si dice grafo critico di Ramsey per quello specifico insieme di sottografi. Il rompicapo consiste nel trovare una colorazione per il grafo critico in cui non compaia nessuno dei sottografi. Tale colorazione si dice colorazione critica.

Non ho idea di quale sia il numero di Ramsey per i sei sottografi dati. Il loro grafo completo sarebbe così ampio (contiene centinaia di punti) che non è pensabile giocare su di esso e il relativo rompicapo è troppo difficile per essere risolto al calcolatore. Ciò nonostante i giochi e i rompicapo di Ramsey con grafi completi più piccoli e con matite di soli due colori possono essere molto divertenti.

Il più noto gioco di Ramsey, detto Sim, è stato trattato in questa rubrica nel marzo del 1974. Si gioca sul grafo completo a sei punti ($K_6$) che rappresenta il problema delle sei persone. Non è difficile dimostrare che 6 è il numero di Ramsey per i due sottografi seguenti:

1. Rosso: triangolo ($K_3$)
2. Blu: triangolo ($K_3$)

Nella teoria classica di Ramsey si è soliti usare solo numeri per i grafi completi; possiamo quindi esprimere il risultato precedente con questa notazione concisa: $R(3,3)=6$. Questo significa che $R$, il numero di Ramsey per il più piccolo grafo completo che comporti un triangolo monocromatico (tutto rosso o tutto blu) quando il grafo è a due colori, è 6. Quindi, se ci sono due giocatori che colorano il $K_6$, a turno, di rosso e di blu un giocatore perderà sicuramente completando un triangolo del suo colore. Il facile rompicapo corrispondente è quello di colorare a due colori il grafo critico $K_5$ in modo che non compaia nessun triangolo monocromatico.

Risulta che, quando $K_6$ viene colorato con due colori, ci sono necessariamente almeno due triangoli monocromatici. (Se ce ne sono esattamente due e sono di colore opposto, formano un farfallino.) Questo pone una domanda interessante: in un grafo completo a $n$ punti a due colori, quanti triangoli monocromatici ci sono necessariamente? A.W. Goodman fu il primo a rispondere a questa doman-


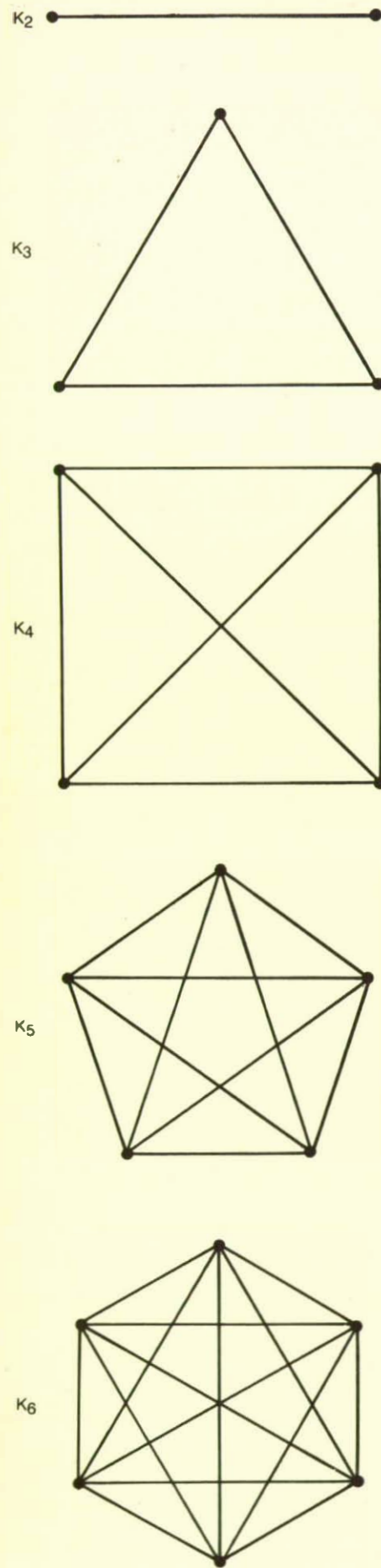

*Colorazione critica con tre colori di $K_{16}$, un grafo completo di 16 punti.*

$K_2$

$K_3$

$K_4$

$K_5$

$K_6$

*Grafi completi con un numero di punti compreso tra 2 e 6.*

da in un articolo intitolato *On Sets of Acquaintances and Strangers at Any Party* comparso in «The American Mathematical Monthly» (vol. 66, n. 9, novembre 1959, pagg. 778-783). È meglio spezzare la formula di Goodman in tre parti:

Se $n$ ha la forma $2u$, il numero minimo di triangoli monocromatici è $u(u-1)(u-2)/3$. Se $n$ è $4u+1$, il numero è $2u(u-1)(4u+1)/3$. Se $n$ è $4u+3$, è $2u(u+1)(4u-1)/3$. Quindi per grafi completi da sei fino a 12 punti il numero minimo di triangoli di un solo colore è, rispettivamente, 2, 4, 8, 12, 20, 28 e 40.

Colorazioni casuali con due colori produrranno generalmente un numero superiore di triangoli monocromatici. Quando nel grafo di Ramsey colorato c'è un numero di triangoli esattamente uguale al minimo, si parla di colorazione estremale. Esiste sempre una colorazione estremale in cui i triangoli siano tutti dello stesso colore? (Tali colorazioni sono state dette blu-vuote, a indicare che il numero di triangoli blu è ridotto a zero.) Nel 1961 Leopold Sauvé dimostrò che la risposta è affermativa. Questo suggerisce una nuova classe di rompicapo. Per esempio: si disegni il grafo completo di sette punti. Si può colorarlo in modo che non ci siano triangoli blu e non più di quattro triangoli rossi? Non è facile. (Darò la risposta il mese prossimo.)

Si sa molto poco dei numeri «classici» di Ramsey. Sono il numero di punti del più piccolo grafo completo che comporta un insieme dato di grafi completi più piccoli. Non si conosce nessun procedimento pratico per trovare numeri classici di Ramsey. L'unico algoritmo noto consiste nel provare tutte le colorazioni possibili dei grafi completi, salendo gradino dopo gradino finché non si trova il grafo di Ramsey. La difficoltà del compito tuttavia cresce esponenzialmente e con un incremento così alto che diventa rapidamente impossibile dal punto di vista del calcolo. Ancora meno si sa se vinca il primo o il secondo giocatore in un gioco di Ramsey giocato razionalmente. Il Sim è stato risolto (la vittoria è del secondo giocatore), ma non si sa quasi nulla di giochi di Ramsey che si basino su grafi completi più ampi.

Finora abbiamo considerato solo i giochi di Ramsey del tipo che Harary chiama giochi di fuga. Come egli ha chiarito, sono possibili almeno altri tre tipi di giochi. Per esempio, in un gioco di «conquista» (lungo le linee di Sim), vince il primo giocatore che completa un triangolo monocromatico. Negli altri due giochi il giocatore continua finché non vengono colorate tutte le linee e vince il giocatore che ha il numero maggiore di triangoli del suo colore o quello che ne ha meno. Questi ultimi due giochi sono i più difficili da analizzare, mentre il gioco di conquista è il più facile. In ciò che segue «gioco di Ramsey» indica il gioco di fuga.

Oltre a $R(3,3)$, la base del Sim, si conoscono solo altri cinque numeri classici di Ramsey non banali per colorazioni a due colori:

1. $R(3,4)$ è uguale a 9. Se si colora $K_9$ con due colori si hanno almeno un triangolo rosso ($K_3$) o un tetraedro blu ($K_4$). Non si sa chi vince se si gioca come gioco di Ramsey.
2. $R(3,5)$ è uguale a 14.
3. $R(4,4)$ è uguale a 18. Se $K_{18}$ è colorato con due colori, si ha necessariamente un tetraedro monocromatico ($K_4$). Non è un brutto gioco di Ramsey, anche se la difficoltà di individuare i tetraedri lo rende difficile da giocare. Il grafo e la sua colorazione corrispondono al fatto che in un raggruppamento di 18 persone c'è o un insieme di quattro persone che si conoscono o un insieme di quattro estranei.
4. $R(3,6)$ è uguale a 18. Nello stesso raggruppamento c'è o un insieme di tre conoscenti o sei persone completamente estranee. Come questi due insiemi si colleghino ai due insiemi di tetraedri dell'esempio precedente è una questione interessante.
5. $R(3,7)$ è uguale a 23.

Si noti che l'elenco precedente non comprende $R(5,5)$. Questo accade perché non si conosce ancora il numero di Ramsey per un grafo completo che, se colorato con due colori, comporti un $K_5$ monocromatico. Stefan A. Burr della American Telephone and Telegraph Company, uno dei principali esperti di teoria dei grafi di Ramsey, pensa che $R(5,5)$ non verrà probabilmente mai conosciuto, tanto grande è il salto di complessità. Egli crede che forse lo stesso possa dirsi di $R(4,5)$. In entrambi i casi tuttavia si conoscono i limiti. $R(5,5)$ è incluso tra 38 e 67; $R(4,5)$ è incluso tra 25 e 29.

Si conosce soltanto un altro numero classico di Ramsey, per tre colori. $R(3) = 3$ è banale perché, colorando un triangolo con un solo colore, si ottiene sicuramente un triangolo di un solo colore. Abbiamo già visto che $R(3,3)$ è uguale a 6. $R(3,3,3)$ è uguale a 17. Questo significa che se si colora $K_{17}$ con tre colori, si ottiene almeno un triangolo monocromatico. In realtà se ne ottiene più di uno, ma non si sa esattamente quale sia il numero dei triangoli.

$R(3,3,3) = 17$ venne dimostrato per la prima volta nel 1955. Per il gioco di Ramsey relativo a questo grafo servono matite di tre colori differenti. I giocatori colorano a turno una linea del colore che vogliono, finché un giocatore perde per aver completato un triangolo monocromatico. Chi vince, se entrambi i giocatori fanno le migliori mosse possibili? Non lo sa nessuno. Il rompicapo di Ramsey corrispondente consiste nel colorare con tre colori $K_{16}$, il grafo critico, in modo da non creare nessun triangolo monocromatico. La figura della pagina precedente, tratta da *Graphs and Hypergraphs* dell'esperto francese di teoria dei grafi Claude Berge, illustra una delle due soluzioni differenti. («Differente» è qui usato in un senso combinatorio più profondo di quello con cui si indica la semplice esclusione di rotazione e riflessione.)

Che dire di $R(3,3,3,3)$, il minimo grafo completo che comporta un triangolo di un solo colore quando viene colorato con quattro colori? Non si conosce la soluzione, sebbene sia stato dimostrato che 64 è un confine superiore da John Folkman, un brillante esperto di analisi combinatoria. Il migliore confine inferiore, 51, è stato trovato da Fan Chung, una giovane matematica cinese dei Bell Laboratories che ne dette la dimostrazione nella sua tesi di dottorato.

La teoria di Ramsey può essere generalizzata in molti modi affascinanti. Abbiamo già considerato il modo più ovvio: la ricerca di quelli che vengono chiamati numeri generalizzati di Ramsey per colorazioni a $r$-colori di grafi completi che hanno sottografi diversi da quelli completi. Václav Chvátal e Harary furono i pionieri di questo territorio, mentre Burr vi ha scavato negli ultimi cinque anni. Si consideri il problema di trovare numeri di Ramsey per i grafi completi minimi che comportino una stella monocromatica di $n$ punti. Harary fu il primo a darne una soluzione per una colorazione a due colori. Nel 1973 Burr e J.A. Roberts lo risolsero per un numero qualsiasi di colori.

Un altro problema generalizzato di Ramsey è quello di trovare numeri di Ramsey per colorazioni a due colori di $K_n$ che comportino un numero dato di triangoli monocromatici «disgiunti». (I triangoli sono disgiunti se non hanno punti in comune.) Nel 1975 Burr, Erdös e J.H. Spencer dimostrarono che il numero era $5d$, dove $d$ è il numero di triangoli disgiunti ed è maggiore di due. Il problema è irrisolto per più di due colori.

Il caso generale per le ruote non è risolto neppure per due colori. Il numero di Ramsey per la ruota di quattro punti, il tetraedro, è, come abbiamo visto, 18. La ruota di cinque punti (una ruota con un mozzo centrale e quattro raggi) ha 15 come numero di Ramsey, come è stato recentemente dimostrato dal matematico nigeriano Tim Moon. La ruota a sei punti non è stata risolta anche se si sa che il numero di Ramsey è compreso tra 17 e 20. 20 è il valore ipotizzato. Se è vero, i grafi completi di 17, 18 e 19 punti si possono colorare con due colori senza ottenere una ruota monocromatica a sei punti (cinque raggi). Sarebbe molto interessante trovare una colorazione siffatta per $K_{17}$ e mi farebbe piacere sapere che un lettore l'ha trovata. Se esiste, ha probabilmente un alto grado di simmetria.

La figura della pagina seguente, una preziosa carta fornitaci da Burr e pubblicata qui per la prima volta, cataloga i 113 semplici grafi con non più di sei linee e nessun punto isolato, per i quali si conosce il numero di Ramsey. Si noti che alcuni di questi grafi non sono connessi. In tali casi la struttura completa, tutta rossa o tutta blu, è data dal grafo completo col numero di Ramsey indicato.

Ogni esemplare nella carta di Burr è la base per un gioco e un rompicapo di Ramsey, sebbene risulti che i rompicapo del tipo «trovare colorazioni critiche per i grafi critici» sono meno difficili di quelli del tipo «trovare colorazioni critiche per numeri classici di Ramsey». Si noti che la carta dà sei variazioni del Sim. Una colorazione a due colori di $K_6$ comporta non solo un triangolo monocromatico, ma anche un quadrato, una stella a quattro punti (chiamata talvolta «artiglio»), un cammino di cinque punti, una coppia di cammini disgiunti di due e tre punti (entrambi dello stesso colore), un quadrato con la coda e il semplice «albero» che è al numero 15 sulla carta di Burr. Il triangolo con la coda (8), la stella a cinque punti (12), la croce latina (27) e il pesce (51) potrebbero essere studiati come giochi di Ramsey su $K_7$.

Ronald L. Graham, uno dei maggiori esperti di analisi combinatoria (è a capo del Discrete Mathematics Department ai Bell Laboratories), ha dato molti significativi contributi alla teoria generalizzata di Ramsey. Sarebbe difficile trovare un matematico creativo che assomigli di meno allo stereotipo del matematico. Quando era molto giovane Graham e due suoi amici si esibivano come professionisti del trampolino in un circo col nome di Bouncing Baers. Nel 1968 Graham trovò una soluzione ingegnosa per un problema del tipo di quelli di Ramsey posto da Erdös e András Hajnal. Qual è il più piccolo grafo di tipo qualsiasi non contenente $K_6$ che comporti almeno un triangolo monocromatico quando viene colorato con due colori? La soluzione unica di Graham è il grafo a otto punti che si vede nella figura in alto a pagina 113. La dimostrazione è una semplice *reductio ad absurdum*. Comincia ammettendo che sia possibile una colorazione a due colori che eviti triangoli monocromatici e poi dimostra che questa comporta necessariamente un triangolo siffatto. Almeno due linee partenti dal vertice devono essere, per esempio, grige e la simmetria del grafo ci permette di fare grige le due linee esterne senza perdita di generalità. Gli estremi di queste due linee devono essere congiunti da una linea colorata per evitare la formazione di un triangolo grigio. I lettori possono divertirsi a cercare di completare il ragionamento.

Che fare con problemi analoghi, quan-

| PUNTI | CAMMINI | CICLI | STELLE | RUOTE |
|---|---|---|---|---|
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |
| 5 | | | | |
| 6 | | | | |

*Quattro importanti famiglie di grafi semplici.*

| GRAFO | NUMERO DI RAMSEY | GRAFO | NUMERO DI RAMSEY | GRAFO | NUMERO DI RAMSEY | GRAFO | NUMERO DI RAMSEY | GRAFO | NUMERO DI RAMSEY | GRAFO | NUMERO DI RAMSEY |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 20 | 9 | 39 | 9 | 58 | 11 | 77 | 9 | 96 | 10 |
| 2 | 3 | 21 | 8 | 40 | 9 | 59 | 11 | 78 | 9 | 97 | 11 |
| 3 | 6 | 22 | 7 | 41 | 11 | 60 | 11 | 79 | 9 | 98 | 10 |
| 4 | 5 | 23 | 8 | 42 | 10 | 61 | 11 | 80 | 9 | 99 | 11 |
| 5 | 5 | 24 | 7 | 43 | 11 | 62 | 8 | 81 | 10 | 100 | 11 |
| 6 | 6 | 25 | 8 | 44 | 10 | 63 | 10 | 82 | 9 | 101 | 11 |
| 7 | 6 | 26 | 7 | 45 | 12 | 64 | 11 | 83 | 10 | 102 | 13 |
| 8 | 7 | 27 | 7 | 46 | 14 | 65 | 10 | 84 | 9 | 103 | 12 |
| 9 | 10 | 28 | 8 | 47 | 9 | 66 | 11 | 85 | 10 | 104 | 12 |
| 10 | 18 | 29 | 8 | 48 | 10 | 67 | 11 | 86 | 9 | 105 | 11 |
| 11 | 6 | 30 | 8 | 49 | 9 | 68 | 9 | 87 | 9 | 106 | 12 |
| 12 | 7 | 31 | 10 | 50 | 10 | 69 | 9 | 88 | 9 | 107 | 11 |
| 13 | 6 | 32 | 8 | 51 | 7 | 70 | 9 | 89 | 11 | 108 | 11 |
| 14 | 7 | 33 | 8 | 52 | 10 | 71 | 9 | 90 | 10 | 109 | 14 |
| 15 | 6 | 34 | 10 | 53 | 8 | 72 | 9 | 91 | 11 | 110 | 13 |
| 16 | 9 | 35 | 9 | 54 | 11 | 73 | 9 | 92 | 11 | 111 | 13 |
| 17 | 6 | 36 | 8 | 55 | 8 | 74 | 10 | 93 | 10 | 112 | 15 |
| 18 | 9 | 37 | 9 | 56 | 11 | 75 | 9 | 94 | 11 | 113 | 17 |
| 19 | 9 | 38 | 9 | 57 | 8 | 76 | 9 | 95 | 10 | | |

*Grafi semplici per cui si conosce il numero generalizzato di Ramsey.*



do il sottografo escluso è un grafo completo diverso da $K_6$? La questione non si pone per $K_3$ perché $K_3$ è esso stesso un triangolo. $K_5$ è irrisolto. La soluzione più nota è un grafo di 18 punti descritto nel 1973 in un articolo di Robert W. Irving. $K_4$ è ancora più lontano dall'avere una soluzione. Folkman, in un articolo pubblicato postumo, dimostrò che tale grafo di Ramsey esiste, ma la sua costruzione richiede più di $2\uparrow\uparrow\uparrow2^{901}$ punti. Questo è un numero così mostruoso che non c'è modo di esprimerlo senza usare una speciale notazione con frecce. Tale notazione è stata introdotta da Donald E. Knuth nel suo articolo *Mathematics and Computer Science: Coping with Finiteness* apparso in «Science» il 17 dicembre del 1976.

Immaginatevi l'universo riempito con sfere della grandezza di elettroni in stretto contatto. Il numero totale di tali sfere è incredibilmente minore di quello dei punti occorrenti nel grafo di Folkman. Erdös offre cento dollari a chi trova un grafo per questo problema che abbia meno di un milione di punti.

Il grafo di Folkman mostra drammaticamente come un problema di Ramsey possa essere enormemente difficile, anche quando l'enunciato del problema non nomina grafi con più di quattro punti. Ma, come amava dire Al Jolson, il meglio deve ancora venire. Graham ha trovato un esempio che è ancora più sconvolgente.

Si consideri un cubo con linee che congiungono ogni coppia di vertici. Il risultato è un grafo completo di otto punti con in più una struttura geometrica euclidea. Si immagini di colorare le linee di questo $K_8$ spaziale di rosso e di blu. È possibile farlo in modo che non risulti nessun $K_4$ monocromatico *che giaccia su un piano*? La risposta è positiva e ciò si può fare facilmente.

Generalizziamo la cosa a cubi $n$-dimensionali. Un ipercubo ha $2^n$ vertici. Anche sull'ipercubo quadridimensionale è possibile colorare con due colori le linee del grafo completo di $2^4$ ossia 16 punti in modo che non risulti nessun grafo planare completo di quattro punti di un solo colore. Lo stesso si può fare con il $2^5$ ipercubo di 32 punti. Questo suggerisce il seguente problema euclideo di Ramsey: qual è la più piccola dimensione di ipercubo per cui, se si colorano con due colori le linee che congiungono tutte le coppie di vertici, si otterrà necessariamente un $K_4$ planare di un solo colore? Il teorema di Ramsey assicura che la domanda ha risposta solo se il $K_4$ non è ristretto a un piano.

Graham e Bruce L. Rothschild dimostrarono per primi, in una profonda generalizzazione del teorema di Ramsey che trovarono nel 1970, l'esistenza di una risposta per un $K_4$ monocromatico planare. Trovare realmente il numero è, tuttavia, un'altra cosa. In una dimostrazione non pubblicata Graham ha di recente stabilito un estremo superiore ma è un limite così grande che detiene il record del più gran numero che sia mai stato usato in una dimostrazione matematica seria.

Per rendere almeno vagamente l'idea della grandezza del numero di Graham, dobbiamo dapprima cercare di spiegare la notazione con le frecce di Knuth. Il numero scritto $3\uparrow3$ è $3\times3\times3=3^3=27$. Il numero $3\uparrow\uparrow3$ sta per l'espressione $3\uparrow(3\uparrow3)$. Dato che $3\uparrow3$ è uguale a 27, possiamo scrivere $3\uparrow\uparrow3$ come $3\uparrow27$ o $3^{27}$. Con una torre pendente di esponenti possiamo scrivere

$$3^{3^3}.$$

La torre ha solo tre livelli, mentre scritta come un numero ordinario è 7 625 597 484 987. È un grosso salto da 27, ma è ancora un numero abbastanza piccolo da poter essere effettivamente scritto.

Quando l'enorme numero $3\uparrow\uparrow\uparrow3=3\uparrow\uparrow(3\uparrow\uparrow3)=3\uparrow\uparrow3^{27}$ viene scritto come torre di 3, ha 7 625 597 484 987 livelli. Sia la torre sia il numero che essa rappresenta sono ora troppo grandi per essere scritti senza una notazione speciale.

Si consideri $3\uparrow\uparrow\uparrow\uparrow3=3\uparrow\uparrow\uparrow(3\uparrow\uparrow\uparrow3)$. All'interno della parentesi si trova il gigantesco numero ottenuto dal calcolo precedente. Non è più possibile indicare in modo semplice l'altezza della torre di tre che esprime $3\uparrow\uparrow\uparrow\uparrow3$. L'altezza è un altro universo separato da $3\uparrow\uparrow\uparrow\uparrow3$. Se spezziamo $3\uparrow\uparrow\uparrow\uparrow3$ in una serie di operazioni con doppia freccia, si ha $3\uparrow\uparrow(3\uparrow\uparrow(3\uparrow\uparrow\ldots(3\uparrow\uparrow3)\ldots))$, in cui il numero di passi da iterare è $3\uparrow\uparrow\uparrow3$. Come dice Knuth, i puntini «evitano molti dettagli». $3\uparrow\uparrow\uparrow\uparrow3$ è oltre ogni immaginazione più grande di $3\uparrow\uparrow\uparrow3$, ma è ancora piccolo rispetto a dove arrivano i numeri finiti, dato che la maggior parte dei numeri finiti è molto più grande.

Ora siamo pronti per indicare il numero di Graham. Esso è rappresentato nella figura in basso. In alto c'è $3\uparrow\uparrow\uparrow\uparrow3$ che dà il numero di frecce del numero immediatamente sotto a esso. Quel numero, a sua volta, dà il numero di frecce sotto di lui. Si continua così per $2^6$ ossia 64 righe. Il numero che sta in fondo è quello che Graham ha dimostrato essere un confine superiore per il problema dell'ipercubo.

*Soluzione di Graham per un problema di Erdös.*

E ora tenetevi forte. Gli esperti della teoria di Ramsey ritengono che il numero di Ramsey per questo problema sia, in realtà, probabilmente 6. Come Stanislaw M. Ulam ha detto molte volte nelle sue lezioni: «Con l'infinito ce la sbrighiamo in fretta; il finito può richiedere un po' più di tempo».

Il problema del mese scorso era quello di indovinare la regola segreta che determinava la disposizione finale delle carte in una mano del gioco dell'Eleusi. La regola era: «Se l'ultima carta è minore dell'ultima carta valida giocata, si giochi una carta più alta dell'ultima, altrimenti se ne giochi una più bassa. La prima carta giocata è corretta a meno che sia uguale alla carta iniziale».

3 ↑ ↑ ↑ ↑ 3

3 ↑ ↑ ..... ↑ ↑ 3

3 ↑ ↑ ............ ↑ ↑ 3

⋮

3 ↑ ↑ ........................ ↑ ↑ 3

64 FILE

*Confine superiore trovato da Graham per la soluzione di un problema euclideo di Ramsey.*